

Open Research Online

The Open University's repository of research publications and other research outputs

Directed evolution of Diels-Alderase ribozymes using in vitro compartmentalisation.

Thesis

How to cite:

Agresti, Jeremy Jon (2006). Directed evolution of Diels-Alderase ribozymes using in vitro compartmentalisation. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 2006 Jeremy Jon Agresti



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Version of Record

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.21954/ou.ro.0000f657>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

**Directed evolution of Diels-Alderase ribozymes using
in vitro compartmentalisation**

A thesis submitted to The Open University, UK
for the degree of Doctor of Philosophy in Biochemistry

By

Jeremy Jon Agresti, B.S., M.S.

Sponsoring establishment:
Medical Research Council Laboratory of Molecular Biology,
Division of Protein and Nucleic Acid Chemistry,
Cambridge

Submitted 27 June, 2005

DATE OF SUBMISSION 27 JUNE 2005
DATE OF AWARD 24 FEBRUARY 2006

ProQuest Number: 13917292

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 13917292

Published by ProQuest LLC (2019). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Abstract

Darwinian evolution in nature has produced many remarkable biocatalysts. By emulating the same processes in the laboratory there is great potential to produce catalysts that rival natural enzymes for efficiency, and are tailor made to the needs of science and industry. In addition, it may be possible to gain a deeper understanding of the evolutionary process itself. *In vitro* compartmentalization (IVC) has previously been used to evolve protein enzymes. In this thesis, I demonstrate how IVC can be applied to select RNA enzymes (ribozymes) for a property that has until now been unselectable: true intermolecular catalysis. Libraries containing up to 10^{11} ribozyme genes are compartmentalized in the aqueous droplets of a water-in-oil emulsion, such that most droplets contain no more than one gene, and transcribed in situ. By co-encapsulating the gene, RNA, and the substrates/products of the catalysed reaction, ribozymes can be selected for all enzymatic properties: substrate recognition, product formation, rate acceleration and turnover. Starting from a large random library of RNAs, a previous study using SELEX, which is based on selection for single-turnover, intramolecular reactions, yielded Diels-Alderase ribozymes which all contained the same catalytic motif. Selecting exactly the same library using IVC, in addition to ribozymes with the motif found in the SELEX study, it was possible to select Diels-Alderase ribozymes with a completely novel catalytic motif. All the selected ribozymes were capable of catalysing the reaction between dienes and dienophiles in solution in a truly bimolecular fashion and with multiple turnover. Interestingly, the catalytic properties of all the selected ribozymes were quite similar. The ribozymes are strongly product inhibited consistent with the Diels-Alder transition state closely resembling the product. More efficient Diels-Alderases may also need to catalyse a second reaction that transforms the product and prevents product inhibition

Table of Contents

Abstract.....	3
Table of Contents.....	5
Table of figures	8
Chapter 1	9
Introduction	9
Directed evolution of biocatalysts	11
The introduction of variation	11
Linking genotype to phenotype	18
Selecting for catalytic activity.....	20
Ribozymes	32
The Diels-Alder reaction.....	34
Kinetic control of the Diels-Alder reaction	36
Chapter 2	42
Methods	42
Reagents.....	42
Ribozyme constructs.....	43
49mer and Δ 49mer	43
69mer and Δ 69mer	46
DNA for selections.....	46
Synthesis of substrates and products.....	46
9-Anthracenylmethyl hexaethylene glycol (AHEG).....	46
Anthracene-HEG phosphoramidite	48
Anthracene-labelled oligonucleotides	48
Diels-Alder adduct.....	48
Library Construction	49
157-mer library	49
69mer libraries	50

Ribozyme selections	50
Compartmentalized transcription	50
Breaking emulsions	51
Gene capture	52
PCR amplification of enriched genes	52
Assaying Diels-Alderase activity of selected libraries.....	53
Characterization of selected clones.....	53
Large scale RNA preparation	54
Structural probing of selected clones	55
Labelling RNA	55
Inline attack probing	56
Probing by partial nuclease digestion.....	56
Kinetic Characterization of selected clones	57
Determination of kinetic constants	57
Determination of product inhibition constants	58
Chapter 3.....	61
Optimisation and model selections	61
Introduction.....	61
Results and discussion.....	66
Optimisation of emulsified transcription	66
Optimisation of emulsified enrichment.....	74
Model selections	80
Chapter 4.....	85
Selection of libraries	85
69mer libraries	85
157mer library	88
Characterization of selected clones.....	98
Chapter 5.....	104
Discussion.....	104
69mer selections.....	107
157mer selections.....	109

References.....	120
-----------------	-----

Table of figures

Figure 1. The catalytic cycle.	26
Figure 2. Two IVC strategies.	30
Figure 3 Simplified mechanism of the Diels-Alder reaction in terms of frontier molecular orbital (FMO) theory. The addition of an unsymmetrical diene and dienophile is shown with the four possible stereoisomers. Electrons of the same phase (color) are able to interact and form the new c-c bonds.	35
Figure 4 Selection of Diels-Alderase ribozymes using IVC.	40
Figure 5 Constructs for cloning and selection.	45
Figure 6. ^1H NMR of AHEG	47
Figure 7. ^1H NMR of rac4.	49
Figure 8 Synthesis of substrates and products	59
Figure 9 Linearity of Gelstar-stained PAGE gel for quantitation.	67
Figure 10 Transcription in the presence of DTT and Maleimide	69
Figure 11 Effect of Mg^{2+} on transcription yield.	71
Figure 12 Effect of NaCl concentration of transcriptional yield	72
Figure 13 Transcription under different emulsification conditions	74
Figure 14 Ribozyme activity on double-stranded DNA.	76
Figure 15 Mock selection of 49mer ribozyme.	78
Figure 16 Effect of buffer Mg^{2+} and temperature on enrichment.	79
Figure 17 Model selection of 69mer ribozyme by IVC.	83
Figure 18 Libraries of the IL region of the 69mer ribozyme.	86
Figure 19 Selection of 157mer libraries by IVC.	89
Figure 20 Secondary structure of ribozymes.	91
Figure 21 Secondary structure probing.	92
Figure 22 Secondary structure probing.	93
Figure 23 Diels-Alderase sequences selected by IVC.	97
Figure 24 Kinetic plots of IVC-selected clones.	99
Figure 25 Kinetic plots of IVC-selected clones.	100
Figure 26 Secondary kinetic plots of IVC-selected clones.	101
Figure 27 Product Inhibition plots of IVC-selected clones.	103
Figure 28 Kinetic parameters of natural and artificial Diels-Alderases.	118

Chapter 1

Introduction

All enzymes that exist in nature are the product of Darwinian evolution by natural selection. Over billions of years the evolutionary process has produced an astounding array of exquisite catalysts that can accelerate chemical reactions as much as 10^{17} -fold, and do so under environmental conditions that are compatible with life; enzymatic transformations are carried out at ambient temperatures, pressures, and near neutral pH in an aqueous environment. Nearly all existing enzymes are proteins, however since the 1980s, several naturally occurring RNA enzymes have been discovered, giving evidence for the existence of a precursor to our current biological system where nucleic acids were the sole information-carrying and catalytic molecule, the so-called “RNA world” (1).

Darwinian evolution can also be reproduced in the laboratory, which has led to the development of a multitude of man-made binding molecules and catalysts, both proteins and nucleic acids. Many investigators have chosen *in vitro* evolution methods to study the structure and function of existing enzymes or to create new enzymes with novel catalytic function. Others have used these methods to study the nature of the evolutionary process itself, using the controlled setting of the laboratory to probe selection conditions and evolutionary intermediates more precisely than can ever be done by studying the natural process.

This thesis describes the first application of a recently described technique, *in vitro* compartmentalisation (IVC) (2) to the directed evolution of RNA catalysts (ribozymes). In particular, I use the Diels-Alderase ribozyme previously described by Andres Jäschke and Burkhard Seelig (3) as a model to demonstrate the directed evolution of a ribozyme by IVC. The introduction covers major aspects of the directed evolution of both protein and nucleic acid catalysts since so many methods and considerations are relevant to both classes of molecules. In addition, some background on the Diels-Alder [4+2] cycloaddition reaction that the ribozymes catalyse is given. Chapter 2 gives experimental details and methods. Chapter 3 begins with a detailed study of *in vitro* transcription and RNA catalysis in emulsions, which leads to the demonstration of transcription and subsequent enrichment of active Diels-Alderase ribozymes in a model experiment. Chapter 4 demonstrates how random libraries of potential Diels-Alderase genes can be selected for true catalysis using IVC. Finally, In chapter 5, I discuss the results of the selections and how they may give some insight into the potential of Diels-Alder catalysis.

Directed evolution of biocatalysts

The introduction of variation

The raw material for the evolutionary process is a repertoire of genes that vary slightly from each other. In nature the repertoire could be a population of organisms, each very similar to one another, but still unique. In the laboratory the “populations” are groups of similar genes, each encoding a slightly different protein or nucleic acid, and are known as libraries. Given a certain length gene, the total set of all possible sequences is known as sequence space.

Sequence space is vast. For example to have a complete library of all possible variants of a 157-nucleotide RNA would require 4^{157} (3.3×10^{94}) different molecules. The complete repertoire would weigh 2.8×10^{72} kg, about 10^{20} -fold more than the estimated mass of the entire universe. Amino-acid sequence space is even larger, with 20 possible amino acids at each position, the total protein sequence space is 20^n for a protein of length n . Clearly, only a miniscule fraction of possible proteins and nucleic acids have ever, and will ever be explored by natural or laboratory evolution. The incredible variation of catalytic function that exists in nature is that much more amazing when we consider that this is only a fraction of what is possible.

For nucleic acid evolution experiments, it is possible to use library sizes of up to $\sim 10^{16}$ variants. For proteins, libraries of up to 10^{10} have been selected for binding, but selections for catalysis usually begin with much smaller repertoires due to the nature of currently available selection methods.

It is logical that the more of sequence space an experiment can explore, the greater its chance of success. For example, it has been shown in both simulations and experiments of selection for antibodies, that the probability of finding a molecule with a given affinity to a ligand increases with increasing library size. In addition, the probability of finding a higher affinity binder is higher in a larger library (4-6). The same relative relationship probably holds true for catalytic molecules, and it has been shown that catalysts are about as common as binders (aptamers) in nucleic acid libraries (7).

One begins an evolution experiment by screening the initial library with a relatively low selection threshold so that sequences with even modest activity will be selected. In subsequent rounds, more variation can be introduced into the selected sequences to explore related areas of sequence space not represented in the initial library. The evolutionary process can proceed for many iterations of selection and modification until an optimal catalyst is selected.

Libraries can be created using dozens of well-described methods. There are arguments to be made for or against all of them, and it is common to combine two or several methods in one evolution experiment.

Random synthesis

In nucleic acid evolution, often a completely random stretch of DNA of between 50 and 150 nucleotides is synthesised chemically. The efficiency of synthesis drops greatly with oligonucleotides greater than about 100 nt, so two shorter random sections can be synthesised and ligated together to give longer genes. A constant region large enough to contain restriction sites or a primer binding site is often put

on one or both ends to allow for manipulation and cloning. Random synthesis is the most naïve method possible for making a library; one only makes the assumption that the length of the gene will provide enough opportunities for the resulting nucleic acid to fold into a structure that can facilitate the reaction or binding property.

Random PCR mutagenesis

It is also possible to use PCR to add variation to a gene or selected library.

Thermophilic DNA polymerases can be made to be even more error prone by adjusting the relative concentrations of dNTP, and by adding Mn^{2+} in addition to Mg^{2+} . The rate of mutation can be precisely controlled, but the limitation of this method is that certain types of mutations are more common than others. For example, transition mutations (purine to purine or pyrimidine to pyrimidine) are much more common than transversions (purine to pyrimidine or pyrimidine to purine), which skews the spectrum of mutations observed. A variation on error-prone PCR using the universal-pairing base analogs 6-(2-deoxy-b-D-ribofuranosyl)-3,4-dihydro-8H-pyrimido-[4,5-C][1,2]oxazin-7-one (dPTP) and of 8-oxo-2'-deoxyguanosine (8-oxodGTP) (8) is a superior alternative that results in a much less biased mutation spectrum, and still a precisely controllable mutation rate. These methods are often performed on a single gene, or previously selected library. One limitation of this method is that, unless the mutation rates are extremely high, one is only able to explore the sequence space closely related to the starting sequence.

Saturation mutagenesis

It is also possible to randomize specific residues in a protein or nucleic acid using what is termed saturation mutagenesis. For nucleic acids the desired positions are replaced with N (any base). In proteins this is achieved by replacing codons for the desired amino acid at the gene-level with NNS, where S= G or C. This combination encodes for all 20 amino acids, but only one stop codon. However all codons are not equally represented, and most amino acids are encoded by more than one codon, which makes the library larger than necessary. To avoid these problems another method uses codon-specific trimer phosphoramidites (9) that can be combined in any proportion to give an unbiased sampling of codons. One advantage of this method is that the relative proportions of the various amino acids can be controlled precisely. Also, one can supply a customized subset of codons. For example, only charged or only hydrophobic amino acids could be used. Another advantage is that one can encode all 20 amino acids using just 20 codons, and no stop codons need to be included. Thus, the library is able to encode a larger fraction of useable sequence space for a given repertoire size. Trimer phosphoramidites have recently become commercially available from Glen Research. Saturation mutagenesis requires that yet another level of design is incorporated in the library process, since even though the codons are randomized, the particular residues are chosen by the experimenter.

Recombination

Recombination is another powerful method of creating genetic variation in a library. *In vitro* recombination emulates sexual recombination by allowing the exchange of genetic material between similar “parental” sequences to give rise to

“progeny” with a combination of material from both parents. An important benefit of recombination is that it can bring together independent beneficial mutations from each parent in a more efficient way than random mutagenesis. Methods of recombination can either be based on sequence homology or non-homologous. There are several variations on both general approaches, but only one example from each class is described here.

In the most commonly practiced method of homologous recombination, termed “sexual PCR” or “DNA shuffling” (10), two or more closely related genes are mixed and partially digested with DNase I to create random breaks. The digested fragments are then reassembled via PCR in the absence of additional primers, which allows the fragments themselves to anneal to heterologous strands and act as primers to extend them, thereby creating chimeric genes. A final PCR step with primers flanking the entire gene is performed to create large amounts of double stranded DNA to be used for selection. By controlling the DNase I digestion step it is possible to promote more or fewer crossover events in the resulting chimeras.

Another method of DNA shuffling of homologous sequences, termed the Staggered extension process (StEP) (11), works by annealing primers to the template sequences followed by repeated cycles of denaturation and very short polymerase-catalysed extension. In each cycle the growing fragments anneal to different templates based on sequence complementarity and extend further. Temperature cycling is repeated until full-length sequences form. Due to template

switching, most of the genes contain sequence information from at least two of the different parents.

In recombination experiments, parental genes often encode naturally occurring variants of the same protein from closely related species. In these cases the method is called family shuffling (12), but the general method has also been used to recombine genes that have been selected from randomly mutagenized libraries in directed evolution experiments (13).

Although DNA shuffling has proven to be a valuable tool for creating genetic variation in directed evolution experiments, the reliance on high levels of DNA sequence homology limits the potential scope for recombination partners. In order to be successful the parental genes must have a large amount of DNA sequence identity. Although genes with DNA identities of as low as 56% have been successfully shuffled (14), there is normally a need for 70-80% identity. Many proteins have similar structures that share very little sequence identity at either the amino acid or protein level, and it has been shown that rearrangement of structurally-similar components within enzyme superfamilies that share little sequence similarity can lead to altered function (15). The development of a method of non-homologous recombination, which allows one to recombine these structurally-similar but sequence-divergent genes would be a significant advance.

Liu and co-workers have developed a particularly elegant method, termed non-homologous random recombination (NRR) and applied it to the directed evolution of both nucleic acid (16) and protein enzymes (17). In NRR, the parental genes are digested with DNase I as in DNA shuffling. After digestion, the ends of the

fragments are made blunt with T4 DNA polymerase, and then ligated with T4 DNA ligase in the presence of blunt-ended hairpin adapters. The parental fragments can ligate in any orientation, and in any sequence, even with multiple fragments from the same region being assembled in the same gene or with some fragments missing altogether. Hairpins are used to flank the assembled recombined gene, and titrating the concentration of hairpins in the ligation reaction effectively limits the total length of the reassembled gene. Lower hairpin concentrations will produce larger assemblies, and higher concentrations will produce shorter assemblies. There will always be a distribution of sizes in the resulting ligation, so NRR is an effective way of not only exploring sequence and structural variation, but also length variation in the search for desired catalytic molecules.

Genomic libraries

The environment is a massive store of unexplored genetic and functional variation because only a small fraction ($<<1\%$) of microorganisms can be cultured (18). There are millions of efficient enzymes residing within uncultured microorganisms that are pre-evolved to transform a wide range of substrates. Libraries can be constructed by extracting and randomly cloning gene-sized pieces of DNA from a community of microbes (19). These libraries can be screened by any of the methods described below to search for enzymes that catalyse the desired transformation. In a recent example, DNA was extracted from the gut fauna of termites and moths, and the resulting library was transformed into bacteria, expressed and screened for the presence of carbohydrate-metabolizing xylanases (20). Several novel xylanases were discovered with similar function, but many had

different structure and sequence to known xylanases. The enzyme genes discovered can then be mutagenized or recombined and selected to further optimise the transformation of the desired reaction or perform better under the desired reaction conditions. For example, xylanases recovered by functional screening of genes extracted from cow manure as above were mutagenized by saturation mutagenesis at several amino acid positions and screened for enhanced thermostability. An increase of up to 30°C in denaturation temperature was observed in the selected clones (21).

Linking genotype to phenotype

A fundamental feature in any evolutionary process is the transfer of information from one generation to the next. Since the information itself is rarely under direct selection, there must be a link between the information and the entity that will be selected directly. In biological systems, the information is termed the genotype (a nucleic acid that can be replicated), and the selectable entity is termed the phenotype (a functional trait such as binding or catalytic activity)(22). The directed evolution of nucleic acid binders (aptamers) or catalysts is a special case, in that the genotype and phenotype are inexorably linked as they reside in the same molecule. However, practitioners of laboratory evolution experiments to evolve functional proteins have had to devise a number of methods to link genotype and phenotype.

Display technologies

'Display technologies' physically link genes to the proteins they encode. These include partially *in vivo* methods such as display on bacteriophage, viruses,

bacteria and yeast. And also completely *in vitro* methods such as plasmid-display, ribosome-display, mRNA display, CIS display, DNA display, and microbead display (23-34).

In vivo Compartmentalisation

In nature genotype and phenotype are linked, not by a physical fusion, but by being encapsulated together within cells. DNA and the protein it encodes, along with the products of the catalytic activity all remain together. Compartmentalisation ensures that the beneficial activity of a particular protein will increase the probability of descent to the next generation only for the gene that encodes it. In the laboratory, cellular compartmentalisation is exploited in laboratory experiments by selecting for the activity of heterologous enzymes expressed within bacteria or yeast. *In vivo* selection and screening has been and remains the standard method for the directed evolution of enzymes.

In vitro compartmentalisation

More recently, a method linking genotype and phenotype by *in vitro* compartmentalisation (IVC) has been developed. Instead of cells, genes are encapsulated in the aqueous microdroplets of a water-in-oil emulsion (2). Each microdroplet is between ~0.1 and 2 μm in diameter and ~0.2 - 5 femtolitre in volume (up to about the same size as a bacterial cell), which means that a 50 μl aqueous reaction mixture is dispersed into $\sim 10^{11}$ microdroplets (equivalent to $>10^7$ 1536-well microtitre plates in a single test tube).

Selecting for catalytic activity

Indirect selections

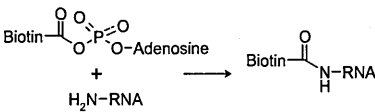
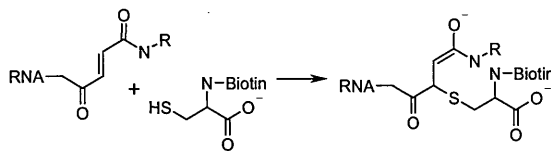
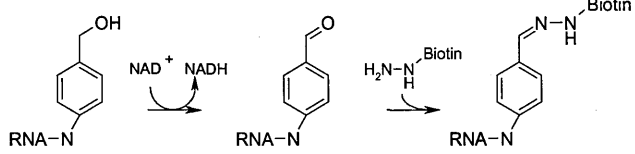
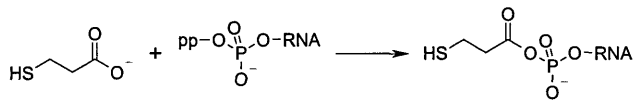
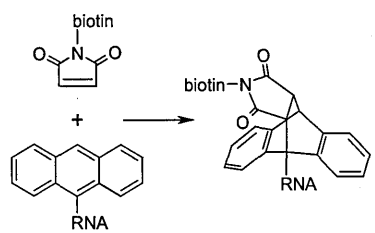
Catalytic function can be evolved by selecting for binding to transition state analogs (TSA) or mechanism-based inhibitors (MBI). Selections can be performed *in vitro* using nucleic acids, antibodies or other proteins. In the case of proteins, display technologies such as those mentioned above are used, and the library is screened for binding to the relevant TSA or MBI. Selections can also be performed for catalytic antibodies *in vivo*. To create catalytic antibodies, the animal is immunized against the relevant TSA (35). While there have been a wide range of catalysts evolved by indirect selection, the method probably has only limited application to evolving efficient enzymes. Rate accelerations are usually low compared to natural enzymes. Furthermore, there are only a limited number of reactions for which TSA or MBI compounds would be available. However, using *in vitro* display technologies such as mRNA display, large libraries of up to 10^{15} variants can be screened, and it may be possible to further evolve these indirectly selected libraries so that they contain potent catalysts if one used a method that was able to select for more of the catalytic (22).

Intramolecular single-turnover selections

Directed evolution experiments for nucleic acid catalysts have largely relied on a selection-amplification technique, termed SELEX (systemic evolution of ligands by exponential enrichment), that was originally developed to evolve nucleic acids with binding activity (aptamers) (36,37). Selections for catalytic molecules using this method are usually referred to in the literature also as SELEX selections, even though the acronym is not accurate.

Most of these selections involve reactions where either bond forming or breaking is accomplished with one reactive group covalently tethered to the repertoire of potential catalysts, while the other is located on a separate substrate or product molecule. Usually, a chemical tag is employed so that a bond forming reaction will attach the tag to the catalyst, and a bond breaking reaction will separate the tag from the catalyst. After allowing time for catalysis, the repertoire is screened by either retaining or rejecting tagged molecules depending on whether the desired reaction was bond forming or breaking respectively. Some examples of the breadth of reactions catalysed by *in vitro* selected RNA are shown in Table 1.

Table 1 Ribozymes accelerating reactions other than nucleic acid-based chemistry.

Type of chemistry (ref)	Selection scheme	Description
Amide synthase (38)		RNA accelerates the reaction of 5' aminated RNA library with AMP-biotin to biotinylate RNA through an amide linkage. Products are separated by streptavidin gel shift. No activity with untethered substrate.
Michael addition (39)		RNA accelerates the reaction of the fumarate-derivitised RNA (Michael acceptor) with the thiol group of biotin cysteine (Michael acceptor). Products are separated on a solid support by streptavidin agarose. No activity with untethered substrate.
NAD ⁺ -dependent oxidation (40)		RNA accelerates the oxidation of RNA-tethered toluene to benzaldehyde with concomitant reduction of NAD ⁺ to NADH. The benzaldehyde is then biotinylated (RNA independently) with biotin hydrazide and separated using streptavidin agarose. Uses untethered NAD ⁺ , but no reaction with untethered toluene.
Acyl transfer (41)		RNA accelerates the acyl transfer of 3-mercaptopropionic acid to the γ phosphate at the 5' end of RNA. The resulting free thiol on the RNA is reacted with pyridine-activated thiopropyl sepharose to covalently link the RNA to the support through a disulfide. Selected RNA was eluted by DTT. No activity on untethered substrate.
Diels-Alder cycloaddition (3)		RNA accelerates the [4+2] cycloaddition reaction between RNA-tethered anthracene, and untethered biotin maleimide. The selected RNA was purified on streptavidin agarose. There is activity with both substrates untethered from the RNA.

Biotin is a common tag utilized in SELEX selections. Successful catalysts that are biotinylated can be captured on a streptavidin support (beads or column) (3,39,40), or bound to free streptavidin and separated on the basis of a mobility shift upon gel electrophoresis (38,40).

The rate accelerations observed in a SELEX experiment cannot be truly classified as being catalysis since, by definition, a catalyst must emerge from the reaction unmodified. However, in some cases it has been possible to modify the selected nucleic acids such that true catalysis is achieved. Often, to achieve intermolecular catalysis the reactive group is coupled to an oligonucleotide that can associate with the catalyst by virtue of base-pairing interactions (39). While this is technically true catalysis, it is not a particularly informative solution to the problem of evolving a true catalyst. It would be much more desirable to select a catalyst that could recognize a substrate based on affinity for the reactive group itself.

The nature of the SELEX reaction with a single substrate molecule covalently linked to the catalyst leads to several limitations and also a few advantages for selecting novel catalysts. There is selection pressure for substrate affinity for the untethered substrate (if present), rate acceleration, and product formation, but not for substrate affinity of the tethered substrate, product release, or multiple repetitions of the catalytic cycle (turnover).

It has been shown that the expected level of enrichment for a particular catalyst is essentially independent of its affinity for the substrate because the tethered substrate is effectively at an extremely high concentration. All else being equal, a catalyst with extremely high substrate affinity would have at most a 2-fold selective advantage over a catalyst with absolutely no specific binding to the substrate ($K_d=1$) (22). The inability to evolve substrate recognition severely limits the potential of SELEX to produce practical catalysts with desired substrate

specificities. In contrast, most natural enzymes have evolved under intermolecular conditions where there is selective pressure for the enzyme to evolve a substrate affinity (K_M) approximately the equal to the substrate concentration (22).

In an attempt to improve the catalytic efficiency (k_{cat}/K_M) of selected variants, later rounds of SELEX selections are often performed for shorter and shorter time intervals to select catalysts with higher turnover numbers (k_{cat}). Shorter incubations can be effective up to a point, but the maximum selectable k_{cat} is limited by the single-turnover nature of SELEX. The half-life for transformation of the enzyme-substrate complex to product is 0.7 seconds for a k_{cat} of 1 s^{-1} and 0.007 seconds for a k_{cat} of 100 s^{-1} (an average value for natural enzymes). It would be impossible to distinguish between such small amounts of time in a selection. As a result, it is rare for SELEX-evolved catalysts to achieve k_{cat} values of greater than 1 s^{-1} . In contrast, under multiple turnover conditions, the enrichment would be proportional to the amount of product formed by the catalytic reaction, so there would be 100 times more product formed by a 100 s^{-1} catalyst than a 1 s^{-1} catalyst, and the same two catalysts would be easily distinguishable. Furthermore, under multiple turnover conditions it is possible to increase the selective pressure for k_{cat} by reducing the catalyst concentration, requiring more turnover in a given time to reach the threshold of selection.

One advantage of the SELEX method is that enrichment of a particular molecule is mostly dependent on the rate acceleration (k_{cat}/k_{uncat}) it provides, and not on the substrate or catalyst concentration. Therefore, catalysts with extremely low rate accelerations can be selected from a library. The ability to enrich inefficient

catalysts, combined with the ability to select from a large library size of up to $\sim 10^{16}$ variants means that SELEX is an excellent method to evolve catalysts from a completely naïve library, where molecules possessing even low catalytic efficiency would be extremely rare. These libraries enriched for low-efficiency catalysts might make an excellent starting point for further directed evolution experiments that can specifically select for higher catalytic efficiencies.

Despite its limitations, and probably because of these two important advantages, SELEX has produced an astounding array of novel nucleic acid catalysts.

There is an adage in the directed evolution field that is “you get what you select for”. Referring to the fact that the selection pressure placed on a library is the sum of all factors, both intended and unintended. The inverse is even more illustrative for SELEX and other indirect methods of selecting for catalysis, “you won’t get what you don’t select for”. There is no evidence or expectation that these methods will ever evolve a catalyst rivalling a naturally evolved enzyme because of their inability to directly select for high catalytic efficiency.

Intermolecular multiple turnover selections

In order to evolve highly efficient enzymes, it is necessary to put selective pressure on the entire catalytic cycle, from substrate binding and specificity to rate acceleration, product formation, product release, and repetition of the cycle (Figure 1). A selection method must favour catalysts that act upon multiple substrate molecules in free solution for this to be possible. Each catalyst must also transform multiple substrate molecules. This can be achieved by ensuring that the catalyst is

at a lower concentration than the substrate, and that a selectable signal level requires multiple turnover.

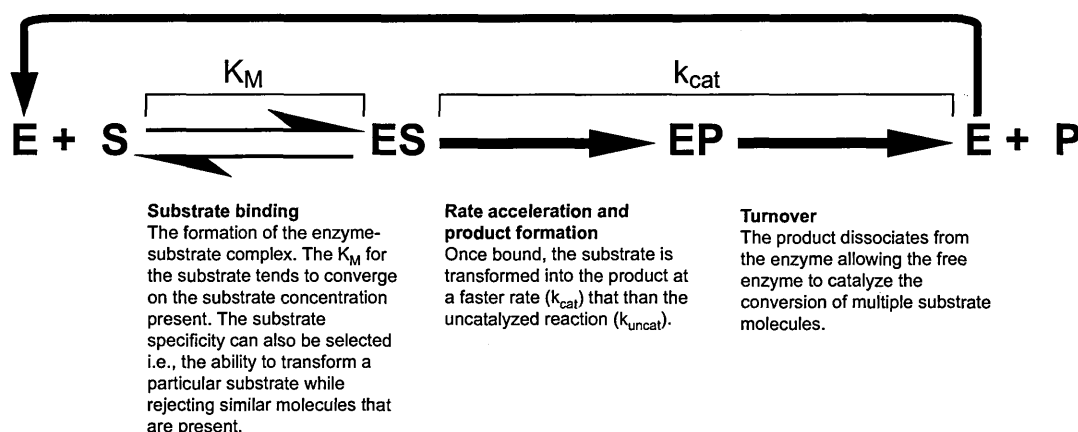


Figure 1. The catalytic cycle.

Intramolecular single-turnover reactions, such as SELEX, are only able to put selective pressure for rate acceleration and product formation, but intermolecular multiple-turnover selections can select for the entire catalytic cycle.

In vivo selections

All natural enzymes have been selected by virtue of intracellular compartmentalization, and most directed evolution experiments to evolve protein enzymes have relied on the same method; heterologous enzymes are expressed within the cytoplasm of yeast or bacterial cells, and the enzyme is allowed to react with untethered substrates that have diffused into the cell. Colonies expressing enzymes with the desired function are usually selected by either linking the desired function to cell survival or by screening for the conversion of a chromogenic or fluorogenic substrate directly on culture plates or in the wells of microtitre plates.

In vivo directed evolution is a powerful and robust method of selection and there are many successful examples of enzymes evolving to act upon different substrates, under different conditions, and also new catalytic reactions discovered by the screening of genomic libraries. However, *in vivo* selection has several limitations that may limit its utility to evolve truly novel catalysts, and to truly explore the catalytic potential of a given reaction.

As discussed above, the probability of finding a protein with a desired catalytic activity increases exponentially with the repertoire size. For *in vivo* selections, the microorganism host must first be transformed with DNA, which places a limit on the maximum library size that can potentially be screened at about 10^8 - 10^{10} . Furthermore, screening even this many individual colonies is a daunting task on plates and Petri dishes. It has been proposed that selecting smaller libraries is not such a serious limitation as long as one does the experiment with multiple cycles of mutation and selection. A potent catalyst would not need to be present in the initial library, and inefficient catalysts that survive the first selection rounds have the opportunity to be improved by further iterations of mutation and selection. This is certainly true, and considering the vastness of sequence space, it is unlikely that a library of any practical size would contain an optimised catalyst at the outset. However screening a larger library to begin with ensures that more possible starting points can be explored.

The selected enzyme must be compatible with the cellular processes of the host organism. The substrates and products of the catalytic activity, and the

environmental conditions of the selection (pH, temperature, salt, presence of solvents, etc.) must not be toxic to cell growth. However, for selections where the desired catalytic activity is not linked to cell survival, catalytic conditions can be adjusted after growth, and so it would be possible to relax some of these restrictions as long as the presence of the enzyme itself was not toxic.

Organisms have a surprising amount of metabolic plasticity, and it is possible that a bacterium will find another solution to grow in the presence of the selection pressure placed upon it than the one planned by the investigator. The enzyme under selection is present within the complex mixture of thousands of other enzymes, and other metabolic pathways can be induced to produce false-positives in the selection.

The next problem is one of dynamic range. In order to be selected, an enzyme must provide enough of an advantage that the cell will grow, but no more than that. Once a certain threshold of catalytic activity is reached, there is no longer pressure to select enzymes with higher catalytic efficiencies. If the threshold for selection is high, then relatively inefficient 'starting points' will not survive to have the opportunity to improve in further rounds. If the threshold is too low, then there is no advantage to being an extremely efficient catalyst over one that is only just adequate for survival

In vitro compartmentalisation

IVC can be used to select nucleic acids and proteins for both binding and catalysis. It has been used to select peptide ligands (32,34,42) and libraries of antibody fragments (Sepp et al., manuscript in preparation). But, so far, the major

application of IVC has been to select protein enzymes (Figure 2). DNA-methyltransferases have been evolved that can efficiently modify novel DNA target sequences (2,43,44). IVC has also been used to select libraries of biotin protein ligase (BirA) (unpublished) and compartmentalisation in emulsions has also been used for the directed evolution of Taq DNA polymerase variants with greater thermostability and expanded substrate range (45,46).

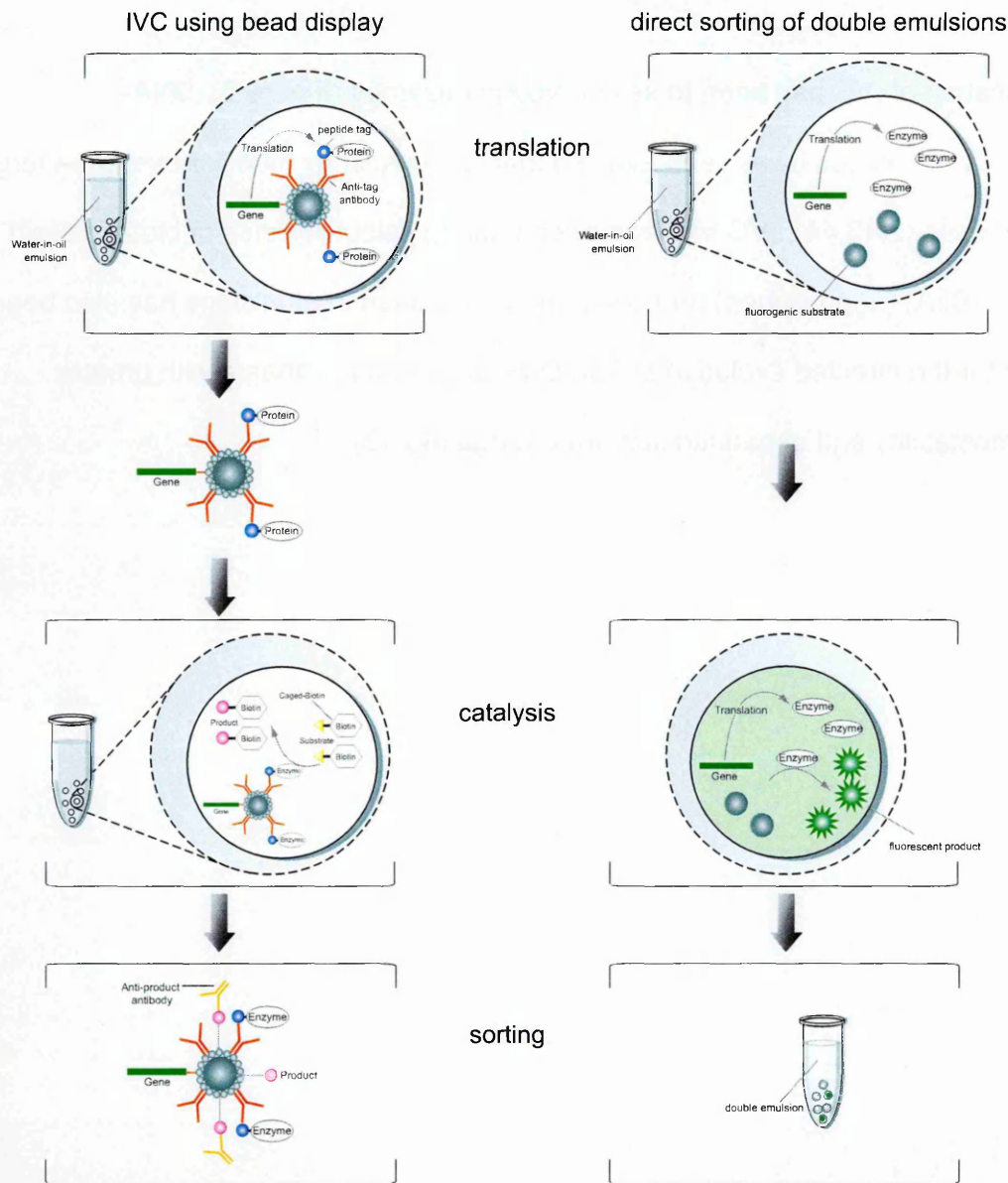


Figure 2. Two IVC strategies. *Left* IVC using bead display as implemented in (47) for the selection of bacterial phosphotriesterase. Single genes attached to streptavidin-coated beads are encapsulated in a water-in-oil emulsion and translated into protein. The protein is captured via a fused epitope tag to an anti-tag antibody on the bead surface. The emulsion is broken, and the beads are emulsified again in the presence of a caged-biotinylated substrate. The substrate is converted into product, the caging group is released from the biotin, and the product molecules bind to the bead. Finally, anti-product antibodies are bound and labelled with fluorescent secondary antibodies, and the beads are sorted by FACS. *Right* IVC using direct sorting of double emulsions as implemented in (48) and (Mastrobattista, in preparation) for the selection of β -galactosidase activity. Single genes are encapsulated and translated within a water-in-oil emulsion in the presence of a fluorogenic substrate. The translated enzymes convert the substrate into fluorescent product. Finally, the water-in-oil emulsion is converted into a water-oil-water double emulsion, and the emulsion drops are sorted directly with FACS.

IVC has many of the advantages of cellular compartmentalization, most notably, the ability to select for rate acceleration on freely dissolved substrates and for multiple turnover (the entire catalytic cycle). The simplicity of an emulsion drop compared to a cell helps IVC to avoid many of the inherent disadvantages of *in vivo* selections. Selection of a particular clone is dependent only on the activity of the enzyme under selection since there are no other metabolic processes present in an emulsion drop to confound the selection. If the protein is expressed in the emulsion drop, there are no other biological processes that it must be compatible with, although if the same emulsion is used for translation and selection, the catalytic conditions must be basically compatible with translation. However some changes can be made to the conditions within the emulsions once they are formed, such as the addition of substrates and inhibitors, changing salt concentrations and pH (47,49) (and in the current study). However, protein and expression and catalysis can be decoupled by combining display technologies with screening for catalytic activity by IVC, which presents few limitations on the conditions that an enzyme can be screened under.

A recent innovation of IVC is to screen water-in-oil-in-water double emulsion droplets directly by fluorescence-activated cell sorting FACS (48) (Figure 2). This method has allowed the evolution of enzymes with β -galactosidase activity from inactive precursors (mastrobattista et. al. in preparation), and the evolution of serum paraoxonase (PON1) mutants with enhanced thiolactonase activity (Aharoni, submitted). Using this method, enzymes can be expressed and screened for the transformation of fluorogenic substrates within the same

compartment, which greatly increases the ease of selection for enzymes where translation and catalysis are compatible.

Ribozymes

All extant naturally occurring nucleic acid catalysts are RNA enzymes (ribozymes). They include the self-splicing group I and group II introns (50,51), the RNA component of RNase P, which generates mature tRNAs by cleaving precursor tRNAs (52), and several self-cleaving RNAs such as the hammerhead, hairpin, HDV (hepatitis delta virus), and VS (*Neurospora* Varkud satellite) motives (53-55). Recently, the RNA component of the large ribosomal subunit has been determined to be a ribozyme, catalysing the peptidyl transferase step of translation (56). Ribosomal RNA is unique among naturally occurring ribozymes because it contains several modified bases. Although modified bases are often used to increase the potential catalytic diversity of RNA in *in vitro* evolution experiments, it is not clear that the modifications of the ribosomal RNA are involved in catalysis. The U2/U6 snRNA complex within the eukaryotic spliceosome might also be a ribozyme (57). In isolation, these two RNA molecules can catalyse a phosphoryl transfer reaction, but the reaction occurs so slowly that it may not be the catalytic activity *in situ*.

In addition to the naturally occurring ribozymes, there has been a large effort to create man-made nucleic acids in the laboratory using the tools of directed evolution. Nucleic acids make excellent candidates for directed evolution because of the innate link between genotype (sequence information) and phenotype (catalytic or binding activity) due to the functional informational molecules being

one and the same. Both ribozymes and DNA enzymes (DNAzymes) (58) have been evolved in the laboratory, and many emulate the catalytic activity of naturally occurring ribozymes. However, others are able to perform catalytic reactions that were previously the sole domain of protein enzymes. The observed catalytic diversity is in spite of the fact that the chemical diversity, and therefore catalytic potential, of nucleic acids is apparently limited in comparison to proteins. There is no nucleic acid base with a pK_a near neutral (as in histidine), a primary amine (as in lysine), a carboxylate (as in aspartate), or a sulfhydryl (as in cysteine). In directed evolution experiments the missing functionality has been supplemented by either using modified bases, or by using unmodified nucleotides. Unmodified nucleotides can recruit small-molecule cofactors or metal ions to increase the chemical diversity.

Many of the man-made ribozymes act upon nucleic acid substrates (reviewed in 59). But focussing on the set of *in vitro* – evolved ribozymes that act upon substrates other than DNA and RNA we see a wide range of catalytic activities. These include amide bond synthesis (38), Michael addition (39), self-alkylation (60,61), acyl transfer (41,62-65), synthesis of enzyme cofactors (66,67), NAD^+ -dependent redox (40), and even carbon-carbon bond formation via the Diels-Alder cycloaddition using both pyridyl-modified (68,69) and unmodified RNA (3).

The evolution of ribozymes that catalyse carbon-carbon bond formation are particularly interesting both to evolutionary biologists studying the origins of life and to synthetic chemists. First, C-C bond forming reactions are necessary for

anabolic processes that must have occurred in the postulated RNA world. Second, transformations catalysed by biomolecules at ambient pressures and temperatures, under near neutral pH, and with a high level of stereo and regiospecificity are particularly attractive to chemists in the fine chemical and pharmaceutical industries.

The Diels-Alder reaction

The Diels-Alder [4+2] cycloaddition reaction between a diene and alkene dienophile is one of the most important and well-studied in organic chemistry. The reaction allows the formation of two simultaneous carbon-carbon bonds and up to four new stereocentres, making it one of the most efficient synthetic methods available to chemists.

Perhaps because of its importance, the Diels-Alder mechanism has also been one of the most hotly debated in history (70). In the current version of the mechanism, the C-C bond formation occurs via concerted mixing of the diene and dienophile π -orbitals (Figure 3) forming a 6-membered ring that proceeds through a 6-membered cyclic (aromatic) transition state. The 6-membered transition state implies that both C-C bonds are formed synchronously in the transition state (71). The highest occupied molecular π -orbital (HOMO) of one species interacts with the lowest unoccupied molecular π -orbital of the other. In other words, the diene HOMO interacts with the dienophile LUMO and vice-versa.

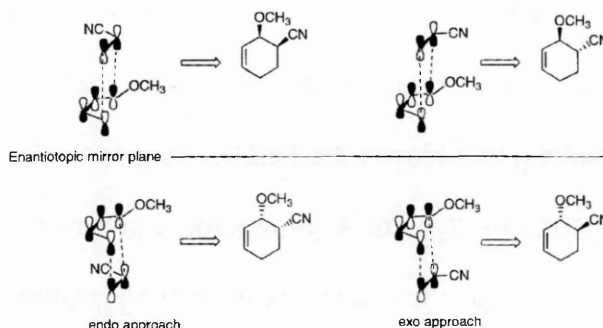


Figure 3 Simplified mechanism of the Diels-Alder reaction in terms of frontier molecular orbital (FMO) theory. The addition of an unsymmetrical diene and dienophile is shown with the four possible stereoisomers. Electrons of the same phase (color) are able to interact and form the new c-c bonds.
Reprinted from (72).

The dienophile usually has π -electron withdrawing groups such as esters, amides, ketones and nitriles, which serve to lower the energy of the dienophile LUMO and increase orbital mixing with the diene HOMO. In a complimentary way, electron-*donating* groups attached to the diene can raise the energy of the diene HOMO, increasing interaction with the dienophile LUMO. Unsubstituted dienes and dienophiles can be made to react under extreme conditions, but most known reaction pairs have substitutions for the reasons described above.

For dienes and dienophiles without additional chiral centers there are two enantiotropic endo and exo transition states possible, leading to four possible stereoisomers (Figure 3). In spontaneous reactions there are only small energy differences driving the formation of the different stereoisomers, and so uncatalysed reactions often form racemic mixtures of products. However, there are opportunities for catalysis of the reaction, which can lead to the asymmetric production of the various isomers.

Kinetic control of the Diels-Alder reaction

Most Diels-Alder reactions proceed through the endo transition state even though the exo path would generally lead to more energetically favourable cyclohexenes, implying that there is something that biases the kinetic processes involved to prevent the formation of the thermodynamically-favoured exo cyclohexene products. In organic solvents the bias is caused by interfering overlap of π -electrons from the electron withdrawing groups on the dienophile (74). In aqueous solution the bias is even more pronounced and reaction rates are often greatly accelerated due to a phenomenon termed the hydrophobic effect. The effect was originally attributed to simple aggregation of the hydrophobic dienes and dienophiles, thereby increasing their effective concentrations and the subsequent reaction rates (75,76), however later evidence has suggested that there may be specific interactions between water molecules and the electron withdrawing groups of the dienophile that may be responsible for the phenomenon (77).

Catalysis of Diels-Alder reactions can also be achieved by the coordination of transition metal ions to the dienophile hetero atoms or to the dienophile. Metals such as Al, B, Ga, Ti, Sn, and Cu have been used as Lewis acids to either raise the energy of the diene HOMO, or (more commonly) lower the energy of the dienophile LUMO. The metal ions have been shown to help to align the substrates into a specific orientation, thereby effecting the regio- and stereospecific bias of the resulting adducts (78). Furthermore, the use of a chiral Lewis acid can even promote enantioselective product formation (79).

Enzymes generally catalyse reactions by stabilizing the structure and charge of the developing transition state. In order for catalysis to be effective, the structures of both the substrates and the product must differ significantly from the transition state. Both the product and the starting substrate must bind to the catalyst less tightly than the transition-state structure for catalysis to occur. The transition state in the Diels–Alder reaction is highly ordered and closely resembles the structure of the product. In other words, an enzyme that was effective at stabilising the Diels–Alder transition state structure would be expected to be inhibited by the product (by tight binding) and turnover (thus, catalysis) would not be possible. Therefore the discovery of biocatalytic Diels–Alder reactions, initially by man-made catalytic antibodies and RNA, and then by natural catalysts was especially exciting since there was the possibility of discovering a new mode of catalysis.

Nature has rarely chosen the Diels–Alder mechanism for the biocatalysis of C–C bond formation despite the utility of the reaction to organic chemists and the potential for catalysis, and instead prefers other mechanisms such as the aldol condensation reaction (80). Although over 100 natural products have been proposed to be Diels–Alder cycloadducts (81), direct evidence from biosynthetic studies with purified or partially purified enzymes is available for only three naturally occurring Diels–Alder reactions (82). The crystal structure of one of these natural Diels–Alderase, macrophomate synthase, has now been elucidated, and confirms that this enzyme forms macrophomate by concerted formation of two carbon–carbon bonds in a Diels–Alder reaction rather than by stepwise carbon–carbon bond formation by a Michael–aldol sequence (83).

However, several antibodies and RNAs which catalyse the Diels-Alder reaction have been generated in the laboratory. The Diels-Alderase antibodies were raised against a hapten that mimicked either the Diels-Alder adduct or the transition state of the desired reactants (reviewed in 84).

A Diels-Alderase ribozyme catalysing the reaction between an acyclic diene and a maleimide-containing biotinylated dienophile (BMCC) were selected by SELEX. The diene was couple through a PEG linker to members of the randomised RNA library. The library was incubated in the presence of the BMCC dienophile, and active catalysts were partitioned by binding to streptavidin followed by gel electrophoresis (68). Pyridyl-modified uridine bases were used to add more chemical functionality to the native RNA. The pyridine modified uridine allows more options for both hydrogen bonding and hydrophobic interactions with the substrates, and also can facilitate the coordination of metal ions that would lead to Lewis acid catalysis. Several active Diels-Alderase ribozymes were selected from a random 100-nucleotide library, the most active, termed DA-22, was found to have complete dependence on Cu^{2+} ions, and was hypothesized to utilize Lewis acid catalysis to enhance the reaction rate. Further selections of mutated libraries of DA-22 (69) have led to increased catalytic efficiency ($k_{\text{cat}}/K_{\text{M}}$) of up to nearly 50-fold, generally due to a decrease in the K_{M} of the BMCC dienophile, although there was a modest increase in k_{cat} for some mutants. None of the selected ribozymes were able to convert untethered substrate, pointing to one of the main limitations of SELEX described above.

Diels-Alderase ribozymes catalysing the reaction between a cyclic diene (anthracene) and a maleimide-containing dienophile (biotin-maleimide) were also selected from a random pool of diene-tethered RNA using SELEX (3). Unmodified RNA was used, and several active clones were isolated from the random 157-nucleotide library, all containing the same basic conserved structural motif. A minimal 49-nucleotide sequence was designed based on the conserved structural motif that was as active as any of the ribozymes selected from the library. The 49mer ribozyme was later shown to act as a true catalyst converting untethered substrates with multiple turnover (3,85). The choice of anthracene as the diene was the likely reason why this ribozyme was able to catalyse the bimolecular reaction. Planar, aromatic, hydrophobic compounds like anthracene have relatively high innate affinity for nucleic acids due to their ability to intercalate. In fact, the dissociation constant (K_d) for several anthryl probes has been shown to be in the same range as the K_M for the anthracene substrate of the selected ribozymes ($\sim 200 \mu\text{M}$) (86).

There is a view from some enzymologists that ribozymes (and deoxyribozymes) may be inherently incapable of generating the same rate enhancements as those achieved by protein enzymes due to the lower chemical complexity of nucleic acids. Others, however, feel that this is not necessarily the case. For example, it has been argued that by using three or four catalytic strategies in parallel a ribozyme or deoxyribozyme should theoretically be able to catalyse RNA cleavage by internal phosphoester transfer at rates at least equivalent to the protein enzyme Ribonuclease A (a nearly 10^{13} -fold rate acceleration) (87). However current

methods of selecting for ribozymes that can rival protein enzymes using intramolecular single turnover reactions are likely to fail since they are unable to select for the entire catalytic cycle from substrate recognition to product formation and release.

With the aim of showing the potential of IVC as a method to select efficient ribozymes, I have chosen to revisit the anthracene-maleimide Diels-Alderase ribozyme first described by Burckhard Seelig and Andres Jaschke (3). In particular, I used a library pre-selected for Diels-Alderase activity by Seelig (3) as a starting point to select specifically for true intermolecular catalysis. I have selected Diels-Alderase ribozymes that catalyse the reaction between dienes and dienophiles in solution in a truly bimolecular fashion and with multiple turnover (Figure 4).

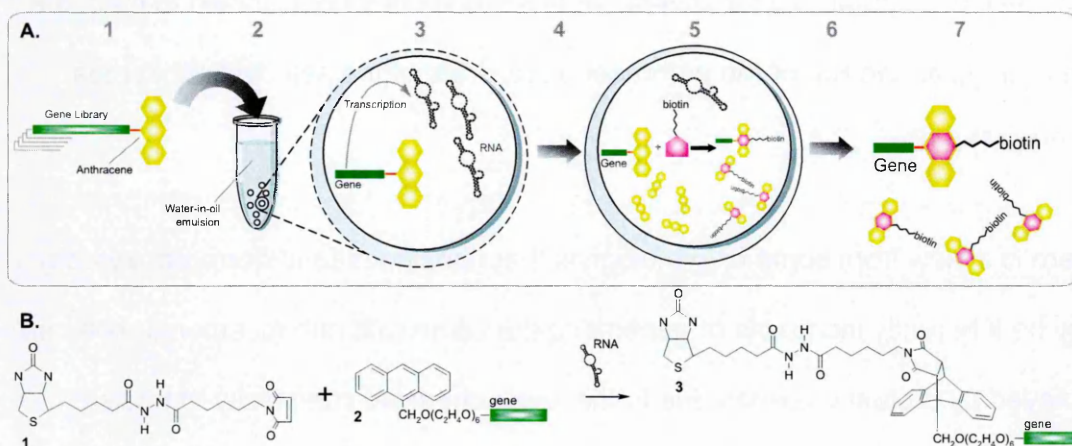


Figure 4 Selection of Diels-Alderase ribozymes using IVC.

Panel A. Schematic diagram of the selection procedure. A repertoire of genes (DNA) encoding ribozymes, each coupled to anthracene through a polyethylene oxide linker, is created (1). The genes are compartmentalized within the aqueous droplets of a water-in-oil emulsion to give, on average, less than one gene per compartment (2). The genes are transcribed giving ~60 RNA molecules per gene (3). Mg^{2+} and biotin-maleimide are allowed to diffuse into the compartments (4). In compartments containing active Diels-Alderase ribozymes, the formation of the cycloadduct by reaction of biotin-maleimide is catalysed, thereby biotinylating genes encoding active ribozymes (5). The emulsion is broken (6), and genes encoding active ribozymes are enriched by binding to

streptavidin-coated magnetic beads (7), and amplified by PCR to allow further rounds of selection. For multiple turnover selections, free anthracene is emulsified with the gene repertoire. **Panel B.** The Diels-Alder cycloaddition of biotin-maleimide (1) and 9-anthracenylmethyl hexaethylene glycol (AHEG) (2) covalently coupled to the gene to generate the adduct (3), thereby biotinylating the gene.

Chapter 2

Methods

Reagents

Restriction enzymes, T4 polynucleotide kinase, and T4 DNA ligase were purchased from New England Biolabs. SuperTaq polymerase was purchased from HT Biotechnology. Calf intestinal alkaline phosphatase was purchased from Roche. T7 RNA polymerase was made in house. All reagents and starting materials for organic synthesis, Biotin-maleimide (**1**) and yeast inorganic pyrophosphatase were purchased from Sigma Aldrich. Pool-5 SELEX DNA "160mer library" was a generous gift of Dr. Andres Jaschke. [Institut für Pharmazie und Molekulare Biotechnologie (IPMB), Ruprecht-Karls-Universität, Heidelberg].

Ribozyme constructs

49mer and Δ 49mer

An insert containing the 49mer Diels-Alderase ribozyme gene (wt49mer) (3), a T7 promoter, and 5' and 3' restriction sites was synthesized using overlapping PCR of three oligonucleotides (Figure 5A): Dazyme-5, Dazyme-mid, and Dazyme-3 (Table 2). Following digestion with *Bgl* II and *Pst* I, this insert was ligated into pIVEX 2.2b Nde (Roche) at the same restriction sites creating pIVEX-49mer (Figure 5D). To create Δ 49mer gene pIVEX-49mer was digested with *Bgl* II and *Sac* I, single stranded ends were removed with mung bean nuclease, and the resulting blunt ends were ligated to recircularize the plasmid creating pIVEX- Δ 49mer in which there is a 40 bp deletion encompassing the T7 promoter and 6 bp from the distal end of the 49mer gene (Figure 5G).

Table 2 Oligonucleotides

Primer	sequence 5'-3'
DA3	GCCGAGTATTGCTCCGAA
DA3andEXT	CAAGGTACCAGGCCGGCGTCGCCGAGTATTGCTCCGAA
DA3EXT	CAAGGTACCAGGCCGGCGTC
DA3EXTPst	GAAGAGCTCGAATTCCTGCAGCAAGGTACCAGGCCGGCGTC
Dazyme-3	CTTGAGGCTGACGTCCGGCTCATAACGAGGCTTCTCGACCGT
Dazyme-5	CATGGATCAAGATCTCGATCCCGCGAAATTCTAATACGACTC
Dazyme-mid	TTTAAGATTATGCTGAGTGATATCCTCGAGCGAAGCCGCTCCGGCACGGTCGAGAAGCCTCGT
J3	GTGGATCCGACCGTGGTG
JA	TCTAATACGACTCACTATAGGAGCTCAGCCTTCACTGC
JBgl	GATCGAGATCTCGATCCCGCGAAATTCTAATACGACTCACTATAG
JPst	ATTGCATTGGCTGCAGGTGGATCCGACCGTGGTGCC
Lib1NWT	CAAGGTACCAGGCCGGCGTCGCCGNNNNNNNTGCGAACANNNNNNNGGCCTCGCCGAAGCGAGCTCCTATAGTGAGTCGTA
Lib2NWT	CAAGGTACCAGGCCGGCGTCGCNNNNNNNNNTGCGAACANNNGGNANNGCCTCGCCGAAGCGAGCTCCTATAGTGAGTCGTA
Lib3NWT	CAAGGTACCAGGCCGGCGTCGCCGANNNNNGCTGCGAACAGCNGGNACGGCCTCGCCGAAGCGAGCTCCTATAGTGAGTCGTA
Lib4NWT	CAAGGTACCAGGCCGGCGTCGCCGNGTATTGCTGCGAACAGNTNNCNCGGCCTCGCCGAAGCGAGCTCCTATAGTGAGTCGTA
LMB2-1	CAGGCGCCATTGCCATT
LMB2-5	CCAGCTGGCGAAAGGGG
LMB2-9	GTAAAACGACGGCCAGTG
LMB2-A	CGCCAGCAACCGCACCTGTGG

* See Figure 5 for location of annealing sites.

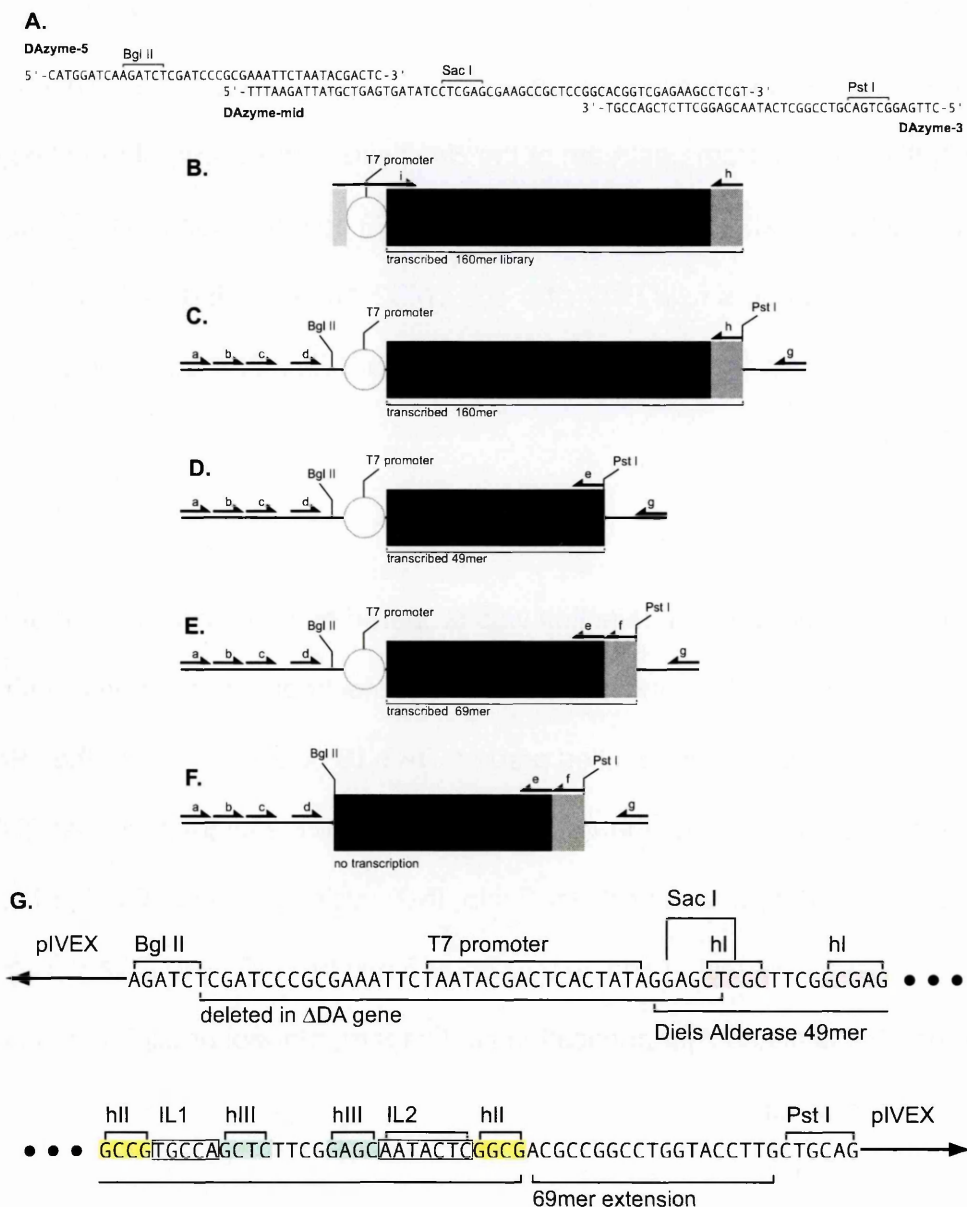


Figure 5 Constructs for cloning and selection.

Panel A, Overlapping PCR diagram for the synthesis of the Diels-Alderase 49mer insert. **Panel B**, PCR product for 157mer library selection. **Panel C**, pIVEX 157mer. **Panel D**, pIVEX 49mer. **Panel E**, pIVEX 69mer. **Panel F**, pIVEX Δ69mer. **Panel G**, A rationally design a minimized 49mer ribozyme that catalyses multiple-turnover Diels-Alder cycloadditions (Diels-Alderase 49mer) (85) was cloned into the vector pIVEX2.2bNde (Roche) downstream of the T7 promoter. The regions encoding helices I, II and III (hI, hII and hIII) and the two halves of the internal loop (IL1 and IL2) of the ribozyme are labeled. We also created a gene that encodes a 69mer ribozyme, which is identical to the 49mer except for a 20-nucleotide extension at the 3' end (69mer extension). The 40 bp deletion in the Δ69mer mutant, encompassing the T7 promoter and the first 6 bp of the ribozyme gene, is also indicated (deleted in Δ69mer) PCR Primer sequences (a-i) are in Table 2.

69mer and Δ 69mer

Using pIVEX-49mer and pIVEX- Δ 49mer as templates, the 69mer and Δ 69mer genes were created by using the primers LMB2-A and DA3andExtPst (Table 2) to PCR amplify a region from upstream of the *Bgl* II site to the 3' end of the ribozyme gene. The DA3andExtPst primer also appends a 20 bp extension to the 3' end of the ribozyme gene and a new *Pst* I site. The LMB2-A/DA3andExtPst PCR product was digested with *Bgl* II and *Pst* I and ligated into pIVEX 2.2b Nde at the same restriction sites.

DNA for selections

Anthracene-labelled DNA for selection was prepared by PCR amplification of ~1 ng of the above plasmids, using SuperTaq (HT Biotechnology), the primer LMB2-5 and either of the anthracene-labelled primers DA3-18 or DA3Ext-18 for the 49mer or 69mer genes respectively (Figure 5) (for oligonucleotide sequences, see Table 2). Reactions were heated to 94°C for 2 min, then cycled 28 times (94°C, 10 s; 50°C, 30 s; 72°C, 30s) with a final extension of 5 min at 72°C. The DNA was gel-purified using a QiaQuick gel-purification kit (Qiagen), ethanol precipitated, and resuspended in water.

Synthesis of substrates and products**9-Anthracenylmethyl hexaethylene glycol (AHEG)**

Hexaethylene glycol (15.9 mmoles) was evaporated twice from 50 ml of dry pyridine to remove traces of water, and then dissolved in 25 ml of dry DMF. To this, 35 mmoles of NaH were added, followed by the slow addition of 15.9 mmoles of 9-chloromethylantracene while stirring. After 24 h 2 ml of glacial acetic acid

were added, and the DMF was evaporated. The crude product was resuspended in 50 ml of dichloromethane (DCM), washed twice with 50 ml of water, twice with 50 ml of 5% sodium bicarbonate, dried over sodium sulphate, and the DCM was evaporated. The singly-substituted product $R_f=0.1$ (TLC DCM:Methanol 19:1) was purified using standard silica chromatography eluting with a gradient from 0-5% methanol in DCM. The relevant fractions were pooled and evaporated yielding 7 mmols (88%) of a viscous amber oil. The identity of the correct product was verified by ^1H -NMR (Figure 6).

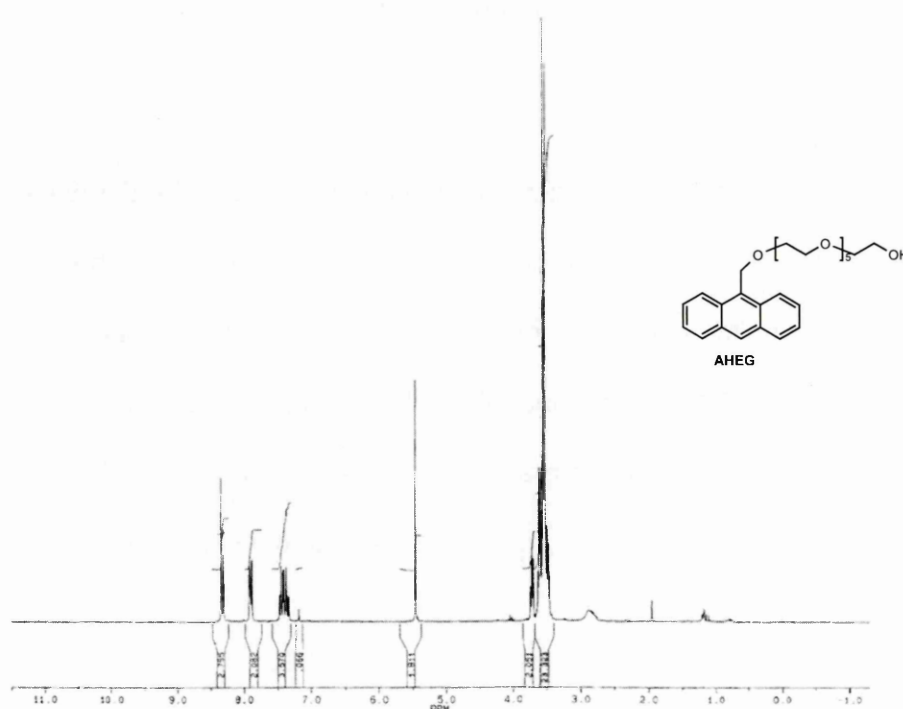


Figure 6. ^1H NMR of AHEG

(CDCl_3 ; 250 MHz) δ 8.31 (s, 1H), 8.18 (d, 2H), 7.88 (d, 2H), 7.18-7.52 (m, 4H), 5.43 (s, 2H), 3.82 (t, 2H), 3.75-3.47 (m, 22H).

Anthracene-HEG phosphoramidite

AHEG (2.5 mmoles) was dissolved in 40 ml of DCM, followed by the addition of 7.5 mmoles of diisopropylethylamine and 3.8 mmoles of 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite. After 1 h the reaction was quenched by adding 100 μ l of methanol, and diluted to 100 ml with ethyl acetate. The crude product was washed twice with 25 ml of 5% sodium bicarbonate, and twice with saturated NaCl, dried over sodium sulfate, and then evaporated to an amber gum. This was purified using standard silica chromatography (9:10:1 ethyl acetate:Hexane:triethylamine), relevant fractions were evaporated to yield 2.1 mmoles (84%) of an amber gum. ^{31}P -NMR was performed to identify the product and showed a single peak as expected (data not shown).

Anthracene-labelled oligonucleotides

Standard solid-phase phosphoramidite chemistry was used to make the DA3Ext-18 and Jaschke3-18 oligonucleotides. Following the incorporation of all of the deoxynucleotides, three units of spacer phosphoramidite-18 (Glenn Research) were added, followed by AHEGP as the final 5' addition, giving a total of 24 polyethylene oxide units between the anthracene and DNA parts of the molecule. The crude oligos were purified by denaturing polyacrylamide gel electrophoresis (PAGE).

Diels-Alder adduct

AHEG (180 μ moles) and 180 μ moles of 6-maleimidocaproic acid (Fluka) were heated for 15 h at 60°C, followed by the addition of an additional 180 μ moles of 6-maleimidocaproic acid. After continuing to stir for a further 24 h, the solvent was evaporated and the crude product purified by silica chromatography eluting with

DCM:MeOH 19:1. This yielded 161 μ moles (90%) of an amber gum. ^1H -NMR was performed confirm the identity of the product (Figure 7).

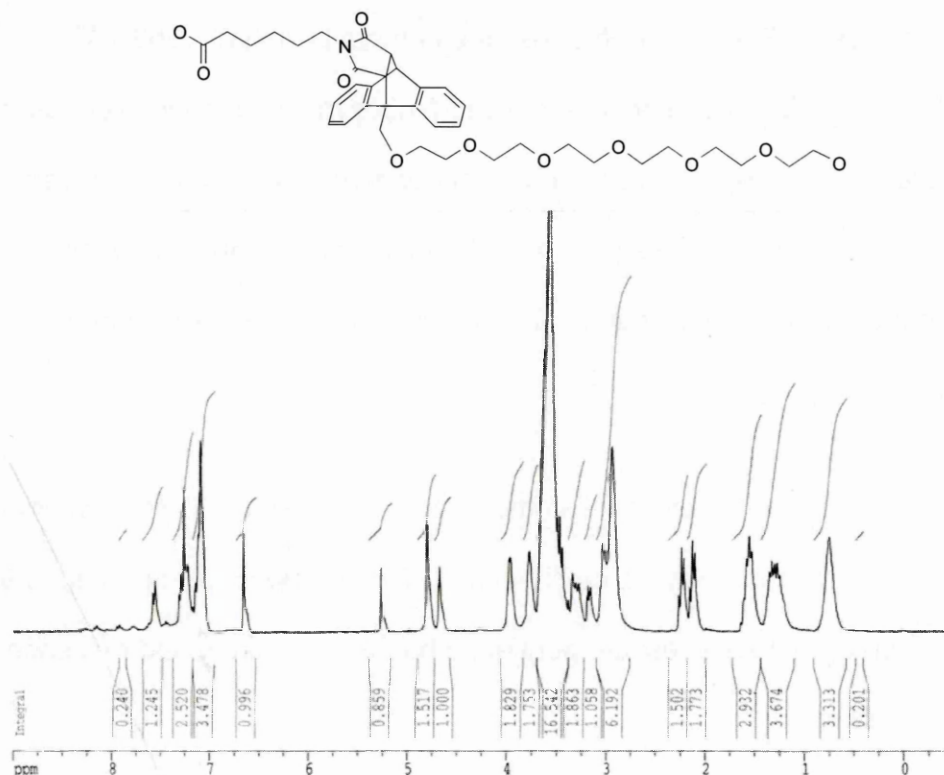


Figure 7. ^1H NMR of rac4.

(CDCl_3 ; 300 MHz) δ 7.58 (d, 1H), 7.27 (dd, 1H), 7.31-7.1 (m, 6H), 4.80 (s, 2H), 4.68 (d, 1H), 3.98 (m, 2H), 3.78 (m, 2H), 3.66-3.44 (m, 20H), 3.34 (d, 1H), 3.03 (dd, 1H), 3.04-2.94 (m, 6H), 2.24 (m, 1H), 2.13 (m, 1H), 1.58 (m, 2H), 1.33 (m, 2H), 0.75 (m, 3H).

Library Construction

157-mer library

1 μ l of a 1000-fold dilution of DNA recovered after the 5th round of SELEX (3) by Burckhard Seelig (a gift of A. Jaschke) was amplified by PCR using 10 μ M each of the primers JA and anthracenylated J3-18 in a 50 μ l reaction. The 179 bp product was gel-purified before selection.

69mer libraries

Libraries 1-4 were amplified by PCR using pIVEX 2.2bNde with no insert as a template. Using a template without the ribozyme gene inserted ensures that there can be no wild-type contamination from the template. The primer upstream of the T7 promoter was LMB2-1 for all libraries, and primers Lib1NWT, Lib2NWT, Lib3NWT or Lib4NWT, which encode the entire ribozyme sequence, were used to amplify libraries 1 to 4 respectively. The library primers encode a compensatory mutation changing the CG base pair to a GC at positions 31/36, hence, all library-derived sequences have this change and are distinguishable from contamination from wild-type sequences.

1 ng of pIVEX 2.2bNde template was amplified by PCR using 10 μ M each primer and SuperTaq polymerase (HT biotechnology). The reactions were heated to 94°C for 2 min, then cycled 28 times, gel-purified, ethanol precipitated and resuspended in water, as above.

To produce anthracene-labelled DNA for selection, approximately 1×10^{11} molecules of each library were PCR amplified using the primers LMB2-5 and the anthracenylated primer DA3Ext-18, gel-purified, ethanol precipitated, and resuspended in water, as for the model selections above.

Ribozyme selections

Compartmentalized transcription

A 50 μ l *in vitro* transcription reaction was prepared on ice containing 40 mM HEPES-KOH pH 7.0, 9 mM MgCl₂, 60 mM NaCl, 5 mM each NTP, 1 U/ μ l RNAsin

(Promega), 1 mg/ml BSA, 0.015 U/ μ l inorganic pyrophosphatase (sigma), 30 μ g/ml T7 RNA polymerase (prepared as in (88)), and 3.3 nM anthracene-labelled ribozyme genes (1×10^{11} genes total) in a 2 ml round-bottom microcentrifuge tube (Eppendorf), and 0.5 ml of ice-cold oil mix [0.5% w/w Triton X-100(Fischer) and 4.5% w/w Span 80 (Fluka) in light mineral oil (Sigma)] were added. The mixture was emulsified by homogenization for 3 min at 25,300 rpm using a Ultra-Turrax T8 Homogenizer (IKA) with a 5 mm diameter dispersing tool.

Transcription was allowed to progress at 37°C for 15 min, and 4 μ l of 3.75 M MgCl_2 dissolved in 50% ethanol were added directly to the emulsion and vortexed vigorously for 10 s. The emulsions were incubated another 15 minutes at ambient temperature, and 1 μ l of 111 mM biotin-maleimide (Sigma) dissolved in DMSO was added directly to the emulsion and vortexed vigorously for 10 s. The reaction was incubated for 15 min at ambient temperature, and 75 μ l of stop buffer were added to the emulsion (100 mM EDTA, 5 mM β -mercaptoethanol, 25 ng/ μ l yeast tRNA).

Breaking emulsions

Emulsions were broken by adding 700 μ l of water-saturated diethyl ether and centrifuging at 25,000 *g* for 1 min. The upper organic phase was removed, and the aqueous phase was extracted first with 350 μ l of neat 2-butanol, then with 700 μ l of water-saturated 2-butanol with removal of the upper organic phase in both extractions. The aqueous phase was extracted a final time with 700 μ l of water-saturated diethyl ether, and centrifuged under reduced pressure for 7 min to

remove traces of ether. At this point the aqueous phase (approximately 100 μ l) was either ethanol precipitated and resuspended in 100 μ l of water, or purified on a G-25 spin column (GE healthcare) to remove excess biotin maleimide.

Gene capture

50 μ l of M-280 streptavidin-coated dynabeads (Dyna) were washed twice in 2xBWT buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 2.0 M NaCl, 0.1% v/v Triton X-100), and resuspended in 100 μ l of 2xBWT buffer. The washed dynabeads were combined with 100 μ l of post-reaction DNA from above in a final concentration of 1x BWT and incubated for 15 min at ambient temperature with gentle rotation to bind DNA. The beads were washed once with 200 μ l of 2xBWT, once with 200 μ l of GHCl Buffer (3M guanidinium hydrochloride, 25 mM sodium phosphate pH 7.2, 1 mM EDTA), once again with 200 μ l 2xBWT, once in 200 μ l of water, and finally resuspended in 25 μ l of water.

PCR amplification of enriched genes

The selected DNA was PCR amplified with SuperTaq from 10 μ l of beads using LMB2-A and either DA3-18 (for 49mer and 69mer selections) or DA3Ext-18 (for 157mer selections) as primers. Reactions were heated to 94°C for 2 min, then cycled 20 times (94°C, 10 s; 50°C, 30 s; 72°C, 45s) with a final extension of 7 min at 72°C. After PCR, DNA was purified using Wizard PCR preps (Promega), and used in further rounds of selection.

Assaying Diels-Alderase activity of selected libraries

DNA amplified from beads after selection was transcribed for 1 h in a 10 μ l reaction in bulk solution (not emulsified) using identical conditions to the emulsified transcriptions described above but with a final concentration of 5 nM. The entire 10 μ l transcription reaction was assayed for activity in 100 μ l 40 mM HEPES-KOH pH 7.5, 80 mM MgCl₂, 60 mM NaCl, containing 100 μ M AHEG and 500 μ M biotin-maleimide in a 96-well plate, by following the reduction in the absorbance at 365 nm over time.

Characterization of selected clones

The selected 157mer library was amplified from beads using the primers JBgI and JPst which append a *Bgl* II and *Pst* I restriction site respectively. The 69mer libraries were PCR amplified using the primers LMB2-A and DA3ExtPst which appends a *Pst* I site. The PCR products were purified with Wizard PCR preps, ethanol precipitated, resuspended in water, and digested with *Bgl* II and *Pst* I. The digested fragments were ligated into pIVEX 2.2bNde digested with the same enzymes, and the ligations were transformed into XL-10 Gold competent cells (Stratagene). Colonies were picked into 384-well plates (Genetix) containing 2xTY, 100 mg/ml ampicillin, 8% glycerol and (75 μ l per well), incubated overnight at 37°C and stored at -70°C.

Bacteria from these plates were transferred using pin replicators (Genetix) into 10 μ l PCR reactions in 384-well plates and amplified using SuperTaq and primers LMB2-A and either J3 or DA3 depending on the library. Reactions were incubated

for 10 minutes at 94°C, and then cycled 30 times (94°C, 30 s; 55°C, 30 s; 72°C, 45s) followed by 7 min at 72°C. One microliter of crude PCR product was used in a 6 µl transcription reaction giving a final DNA concentration of ~20 nM.

Transcription conditions were the same as for the selected pool screen above.

Activity was assayed as above using 5 µl of the transcription reaction in a 55 µl assay in 384-well plates giving a final RNA concentration in the assay of ~3 µM.

Reaction rates of the selected clones were compared to the rate of the wt69mer ribozyme assayed on the same plate, and clones with similar activity to the wt69mer were picked from the 384-well cultures and sequenced.

Large scale RNA preparation

Template DNA for large-scale transcription reactions was prepared in 3.4 ml PCR reactions split into 48 separate 70 µl wells. The primers LMB2-9 and either J3 or DA3 were used to amplify ~200 ng of purified plasmid DNA. Reactions were heated to 94°C for 2 min, then cycled 30 times (94°C, 10 s; 55°C, 40 s; 72°C, 1 min) with a final extension at 72°C for 5 min. Primers were removed by centrifuging in a Centricon YM100 concentrator (Millipore) with two exchanges into water. DNA was extracted once with phenol:chloroform:isoamyl alcohol 25:24:1 (PCI), once with water-saturated diethyl ether, ethanol precipitated, and resuspended in 200 µl of STE buffer (10 mM Tris pH 8.0, 50 mM NaCl, 1 mM EDTA).

The purified PCR products were transcribed in 10 ml with the following conditions: 40 mM HEPES-KOH pH 7.5, 24 mM MgCl₂, 60 mM NaCl, 0.1 mg/ml BSA, 2 mM spermidine, 10 mM DTT, 5 mM each NTP, 0.01% Triton X-100, 1U/ml inorganic

pyrophosphatase, 100 µg/ml T7 RNA polymerase, and 50-70 nM DNA template. After 5 hours of incubation at 37°C, 7 ml of 80% Formamide, 10 mM EDTA loading buffer were added, and the entire 17 ml were loaded on an 8% polyacrylamide, 1xTBE, 8M Urea gel (17x30x0.3 cm). RNA was visualized by UV shadowing, and full-length RNA was excised from the gel and eluted using the crush-soak method overnight in ~100 ml TE buffer (10 mM Tris-HCl pH 7.4, 1 mM EDTA) at ambient temperature. Gel pieces were filtered into several 50 ml conical-bottom centrifuge tubes, 0.1 volumes of 3M sodium acetate pH 5.2 and 2.5 volumes of ethanol were added, and the tubes were centrifuged at 2,750 g for 2 hours. The RNA pellet was washed in 80% ethanol, dried, and resuspended in 500 µl water.

Structural probing of selected clones

Labelling RNA

Approximately 200 pmoles of gel-purified RNA from above were dephosphorylated in 10 µl of 1x calf intestinal pyrophosphatase (CIP) buffer with 4 units of CIP enzyme (Roche) at 50°C for 1 hour in 20 µl. EGTA was added to a final concentration of 10 mM, and the enzyme was heat inactivated at 65°C for 10 min. The reaction was extracted twice with PCI, once with water-saturated diethyl ether, ethanol precipitated, and the pellet resuspended in water. RNA (10 pmoles) was 5' end-labelled with [γ -³²P]-ATP and polynucleotide kinase for 1 hour at 37°C, and the entire reaction was gel-purified on an 8% polyacrylamide, 1x TBE 8M urea gel using the crush-soak method.

Inline attack probing

The procedure is similar to (89). Labelled RNA (50,000 cpm) was heated to 90°C for 1 min and cooled on ice. It was then incubated for 48 hours in a buffer containing 50 mM Tris-HCl pH 8.3, 80 mM MgCl₂, 60 mM NaCl and 1 µg/µl unlabelled yeast total RNA. Reactions were ethanol precipitated, resuspended in a loading buffer containing 10M urea, and analyzed on an 8% polyacrylamide, 1xTBE, 8M urea gel.

Probing by partial nuclease digestion

Labelled RNA (50,000 cpm) was probed for secondary structure using the ribonucleases T1, T2, and V1. RNA in water was heated to 90°C for 1 min and cooled on ice. All reactions were performed in 10 mM Tris-HCl pH 7.4, 10 mM MgCl₂, 50 mM NaCl, 1 µg/µl yeast total RNA. RNA was incubated for 15 min at ambient temperature with 0.1, 0.02, 0.004 U/µl ribonucleases T1 (Ambion) and T2 (Invitrogen), and with 0.01, 0.002, and 0.0004 U/µl ribonuclease V1 (Ambion). All reactions were stopped by ethanol precipitation in the presence of 2 µg/µl glycogen, and resuspended and analyzed by PAGE as above.

The alkaline hydrolysis ladder was produced by incubating 50,000 cpm RNA at 90°C for 7 min in a buffer containing 50 mM sodium carbonate pH 9.2, 1 mM EDTA, and 1 µg/µl yeast RNA. Denaturing T1 digestion was performed by incubating 50,000 cpm RNA at 55°C for 15 min in a buffer containing 20 mM sodium citrate pH 5.0, 1 mM EDTA, and 7 M urea.

Kinetic Characterization of selected clones

Determination of kinetic constants

Purified RNA was heated to 90°C for 2 min and cooled on ice. Activities were measured in 15 mM Tris-HCl pH 7.4, 80 mM MgCl₂, 75 mM NaCl with 50-500 µM AHEG and 2-8 mM MCA. The reactions were performed in 384-well plates in a 70 µl volume with a final RNA concentration of 3 µM. K_m and V^{app} were determined by treating the reaction as pseudo first-order for each substrate and fitting the data from each pair of substrate concentrations to the Michaelis-Menton equation (Equation 1) using the Levenberg-Marquardt algorithm as implemented in KaleidaGraph (Synergy), where $[S_{var}]_0$ is the initial concentration of the substrate being varied.

Equation 1

$$v = \frac{V^{app} [S_{var}]_0}{K_m + [S_{var}]_0}$$

True V_{max} , and therefore k_{cat} was determined by making a secondary plot of the V^{app} values determined above against $[S_{var}]_0$ for each substrate, and fitting the data to Equation 2 in KaleidaGraph, where $[S_{fix}]_0$ is the initial concentration of the fixed substrate.

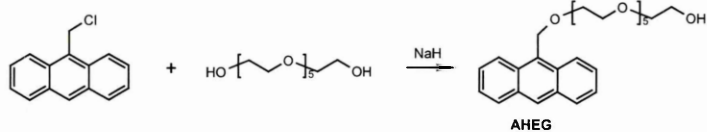
Equation 2

$$V^{app} = \frac{V_{max} [S_{fix}]_0}{K_m + [S_{fix}]_0}$$

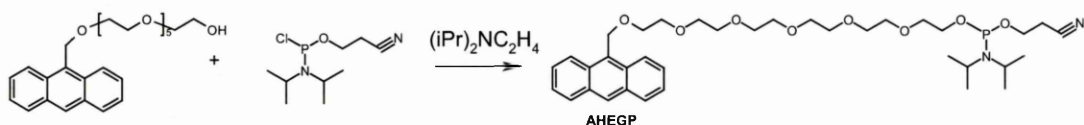
Determination of product inhibition constants

Purified RNA was heated and cooled and initial rates were measured in the same buffer as above using a single concentration of each substrate: 100 μM AHEG and 500 μM MCA. Chemically synthesized racemic Diels-Alder product rac4 (fFigure 8) was added to the reactions from 0 to 1 mM giving a concentration of the more inhibitory *R,R*- enantiomer of between 0 and 500 μM . Initial rates were plotted against the inhibitor concentration and K_i values were determined by fitting the data to a competitive inhibition model assuming substrate binding and complex dissociation are rapid compared to product formation from the termolecular complex and may occur in random order; that the Diels-Alder adduct can only bind to the free enzyme and competes with both diene and dienophile; and that the K_m for each substrate binding to free enzyme is identical to the K_m for the substrate binding to enzyme with the other substrate already bound. (Equation 3) in KaleidaGraph.

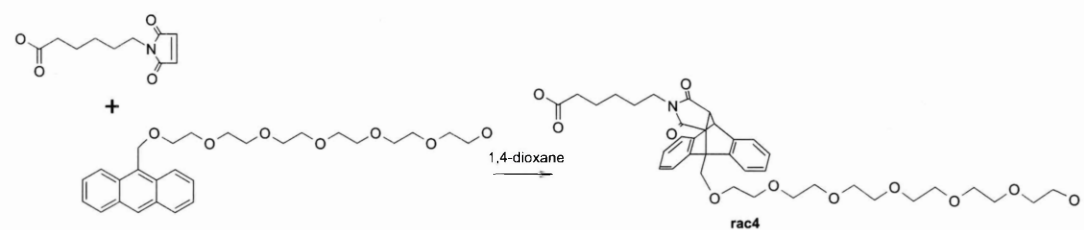
A.



B.



C.

**Figure 8 Synthesis of substrates and products**

A. Williamson ether synthesis of 9-anthracenylmethyl hexaethylene glycol (AHEG) from 9-chloromethylanthracene and hexaethylene glycol **B.** Phosphitylation of AHEG to form the phosphoramidite (AHEGP) **C.** Diels-Alder cycloaddition reaction between maleimidocaproic acid and AHEG to give the racemic adduct **rac4**.

Equation 3

$$v = \frac{k_{\text{cat}}[S_1]_0[S_2]_0[E]_0}{K_{m1}K_{m2}\left(1 + \frac{[i]}{K_i}\right) + K_{m2}[S_1]_0 + K_{m1}[S_2]_0 + [S_1]_0[S_2]_0}$$

Chapter 3

Optimisation and model selections

Introduction

The enrichment factor of a selection is simply the ratio of the amount of product produced by an “active” catalyst vs. the uncatalysed background rate of the reaction ($[P]_{\text{cat}}/[P]_{\text{uncat}}$). If the experiment is designed such that only a small amount of the starting substrate is consumed during the selection, then the ratio of product formed by the catalysed and uncatalysed reactions will be proportional to the ratio of the initial rates ($d[P]/dt$ or v) of product formation ($v_{\text{cat}}/v_{\text{uncat}}$). In the case of the two-substrate reaction catalysed by the Diels-Alderase ribozyme, the uncatalysed second-order rate is given by the equation

Equation 4

$$v_{\text{uncat}} = k_{\text{uncat}} [S_1][S_2]$$

Making the assumption that the reaction occurs via a termolecular complex ($S_1 \bullet S_2 \bullet C$) in which substrate binding is in random order and complex dissociation is more rapid than product formation, Cleland (90) and Dalziel (91) have shown that the rate is approximated by the equation

Equation 5

$$v_{cat} = \frac{k_{cat} [S_1][S_2][C_0]}{K_{M1}K_{M2} + K_{M1}[S_1] + [S_1][S_2]}$$

And therefore

Equation 6

$$\frac{v_{cat}}{v_{uncat}} = \frac{k_{cat}}{k_{uncat}} \cdot \frac{[C_0]}{K_{M1}K_{M2} + K_{M1}[S_1] + [S_1][S_2]} \quad (84)$$

From this, it is clear that in a directed evolution experiment that relies on intermolecular catalysis, such as those using IVC, the enrichment factor in each round of selection is directly proportional to the concentration of the catalyst in the selection (22,84), and it is therefore worthwhile to optimise transcription levels at the outset to maximize the potential enrichment factor.

RNA synthesis by enzymatic transcription initiated from bacteriophage T7, T3, SP6 or similar promoters is a common technique used in molecular biology laboratories for a variety of applications. *In vivo*, such transcription is the standard method by which heterologous proteins are expressed in *E.coli*. In addition, large amounts of RNA can be synthesized easily *in vitro*, and can be used in applications such as: unbiased amplification of cDNA for expression profiling “RNA amplification”, or other isothermal amplification strategies such

as 3SR (92-94). Large amounts of very pure RNA are needed in order to make crystals for structural studies (95,96), and RNA oligonucleotides produced by *in vitro* transcription can be used in gene silencing applications such as RNA interference(97). *In vitro* transcription can also be used to produce randomized populations of RNA molecules for *in vitro* selection experiments to evolve RNAs with binding activity (aptamers)(37) or catalytic activity (ribozymes)(58).

Transcription for all of these applications is done in bulk solution (unemulsified), and with few exceptions, the transcription reaction is allowed to proceed in isolation without the need for conditions to be catalysed for other simultaneous reactions.

The requirements for an *in vitro* transcription reaction are simple. Double-stranded DNA containing the appropriate promoter sequence is incubated with the corresponding RNA polymerase and the four ribonucleotide triphosphates (NTPs) in a buffer providing the appropriate pH, salt, and divalent cation (usually Mg^{2+}) concentrations. RNA synthesis will be initiated at the first nucleotide after the promoter in a 5'-3' direction.

There are no further requirements to produce RNA, but several factors can be added and conditions adjusted to catalyse the yield of the synthesis. For maximum efficiency of initiation the first RNA nucleotide should be a guanosine residue, and the first few nucleotides should be "G-rich". Sequences such as "GGAG" are common at the beginning of the RNA sequence. A reducing agent such as DTT is normally added to prevent oxidation of an active site cysteine residue in the polymerase. A ribonuclease inhibitor such as the human placental

ribonuclease inhibitor (sold commercially as RNAsin by Promega) can increase the net yield by preventing degradation by contaminating nucleases.

There are several interactions between reaction components that need to be catalysed for efficient transcription. The polymerase finds the promoter by binding nonspecifically to double-stranded DNA and tracking until the promoter is found. Because of this general affinity for dsDNA, large amounts of template DNA can sequester active polymerase and cause low transcription yield unless a large amount of polymerase is added to counteract this effect. In the case of emulsified transcription for directed evolution, we have at most a single molecule of DNA at about 1 pg/ μ l (~150 bp template DNA in an ~170 aL compartment), so polymerase sequestration is not an issue.

There is also an interaction between the Mg^{2+} and NTPs in the reaction. NTPs will chelate Mg^{2+} ions nearly stoichiometrically. The NTP/ Mg^{2+} complex is the actual substrate for the polymerase, so there needs to be at least an equimolar ratio of Mg^{2+} /NTP so that all of the NTPs are potential substrates. In fact polymerase transcription efficiency is highly dependent on the amount of free Mg^{2+} (surplus to the NTP concentration) in the reaction. There are conflicting reports in the literature over the relationship between NTP and Mg^{2+} concentrations. While most researchers consider the amount of *free* Mg^{2+} as the important factor, there is one report (98) that adamantly advocates that the *ratio* of NTP to Mg^{2+} is important to consider.

The enzyme inorganic pyrophosphatase (PPi) is also often added to *in vitro* transcription reactions. Pyrophosphate, one of the end products of RNA

synthesis, is thought to be a potent inhibitor of the polymerization reaction. In addition, Pyrophosphate will bind to Mg^{2+} , thereby reducing the amount free in solution. The addition of catalytic amounts of PPI is thought to eliminate these modes of inhibition.

All of the items mentioned above apply to transcription both in bulk solution and in emulsion droplets. But, there are also several factors that will affect emulsified transcription specifically. It has been observed that enzymatic reactions perform very poorly in emulsions unless sufficient amounts of carrier protein are added to the aqueous phase. This is almost certainly due to proteins becoming sequestered at the interface of the emulsion droplets by interaction with the surfactant layer. The addition of bulk protein such as BSA far in excess of the amount of enzyme ensures that only a small amount of enzyme protein will be sequestered.

It is possible to perform enzyme reactions in bulk solution at such catalytically small molar concentrations of enzyme such that there would actually be fewer enzyme molecules than emulsion droplets if the same reaction were emulsified. This segregation effect would cause many compartments to have just a few or no enzyme molecules, thereby reducing the number of "transcription-competent" compartments, and the potential repertoire size screened.

Another set of conditions that needed to be optimised with this particular selection system was that the transcription and catalytic reactions needed to be performed sequentially without breaking the emulsion and purifying the transcribed RNA in between. It has been shown that components can be added

to emulsion droplets by either simple diffusion or by nanodroplet delivery (49), but it is not possible to remove components. Potential interactions with this system are that the maleimide substrate would be at high concentration and would potentially inactivate the T7 polymerase enzyme by reacting with the active site cysteine. DTT is a common addition to keep this Cys residue in the reduced (active) form, but the maleimide substrate would itself become inactive by reacting with the bi-sulfhydryl DTT, which would be in large molar excess.

The following experiments were performed to systematically determine the optimum conditions for emulsified transcription, and then to use these conditions to show the enrichment of active Diels-Alderase ribozymes from a simple mixture of genes encoding active and inactive ribozymes as a prelude to the selection of libraries.

Results and discussion

Optimisation of emulsified transcription

Gel quantitation of RNA on Gelstar-stained gels

A 3-fold dilution series of Diels-Alderase 49mer RNA was loaded on a 12% acrylamide:bisacrylamide 19:1; 1x TBE; 8M urea gel. The gel was post-stained with Gelstar nucleic acid stain (flowgen), a fluorescent dye that is especially efficient at staining single-stranded nucleic acids. The gel was scanned on a Typhoon 8600 multi-mode imager (GE Healthcare) using an excitation of 532 and emission of 520 nm. The wavelengths chosen reflect the laser available on the scanner and not a reverse Stoke's shift of the fluorophore (Gelstar has

Excitation/Emission maxima at 490/530 nm respectively). Densitometry software (Totallab, Nonlinear Dynamics) was used to integrate the volume of the fluorescence in the bands. Figure 9 shows that the signal begins to saturate at about 5 ng/lane, but the quantitation is linear between ~65 pg and 5 ng (a 75-fold range).

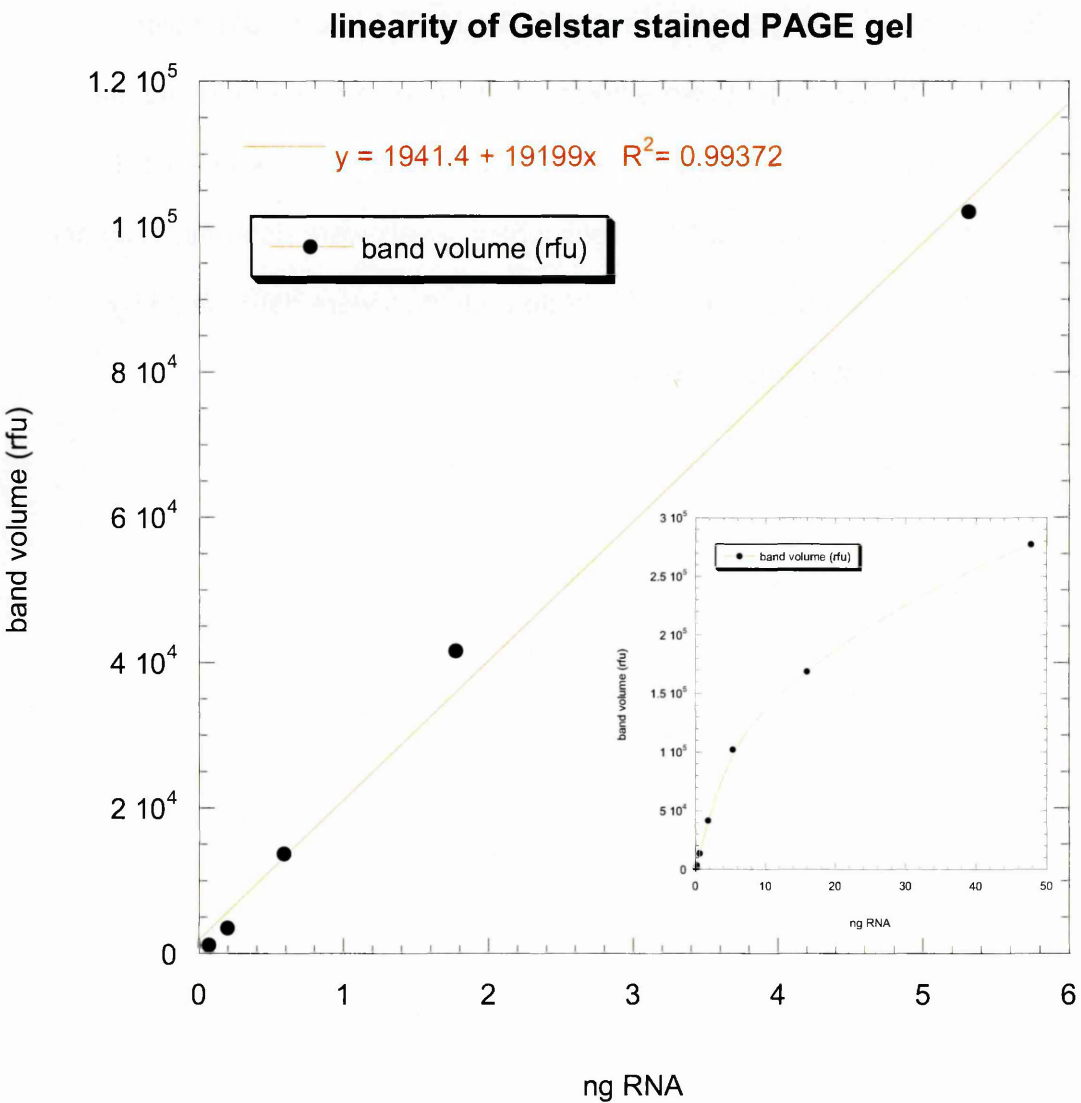


Figure 9 Linearity of Gelstar-stained PAGE gel for quantitation. Gel shows essentially linear signal between 65 pg and 5 ng. Inset shows entire range up to 50 ng,

Effect of DTT and maleimide on transcription

In vitro transcriptions were performed in bulk solution (unemulsified) in 47 mM Tris-HCl pH 8.0 @37°C, 21 mM MgCl₂, 57 mM NaCl, 0.1 mg/ml BSA, 16 mM total NTPs, 4U/μl RNAsin, 0.01% triton X-100, and 10 μg/ml T7 RNA polymerase, and 1 nM template DNA (a PCR product of the 49mer ribozyme amplified with the primers LMB2-A and DA3) (Figure 5). Reactions were performed in the presence or absence of 50 mM DTT and 50 μM biotin-maleimide (BMAL). Aliquots were taken at 10, 30, 60 and 120 min and the reaction stopped by mixing with an equal volume of formamide loading dye (80% formamide, 10 mM EDTA). Transcription levels were determined by gel quantitation as described above. Absolute quantities were determined by comparison with standard lanes (Figure 10)

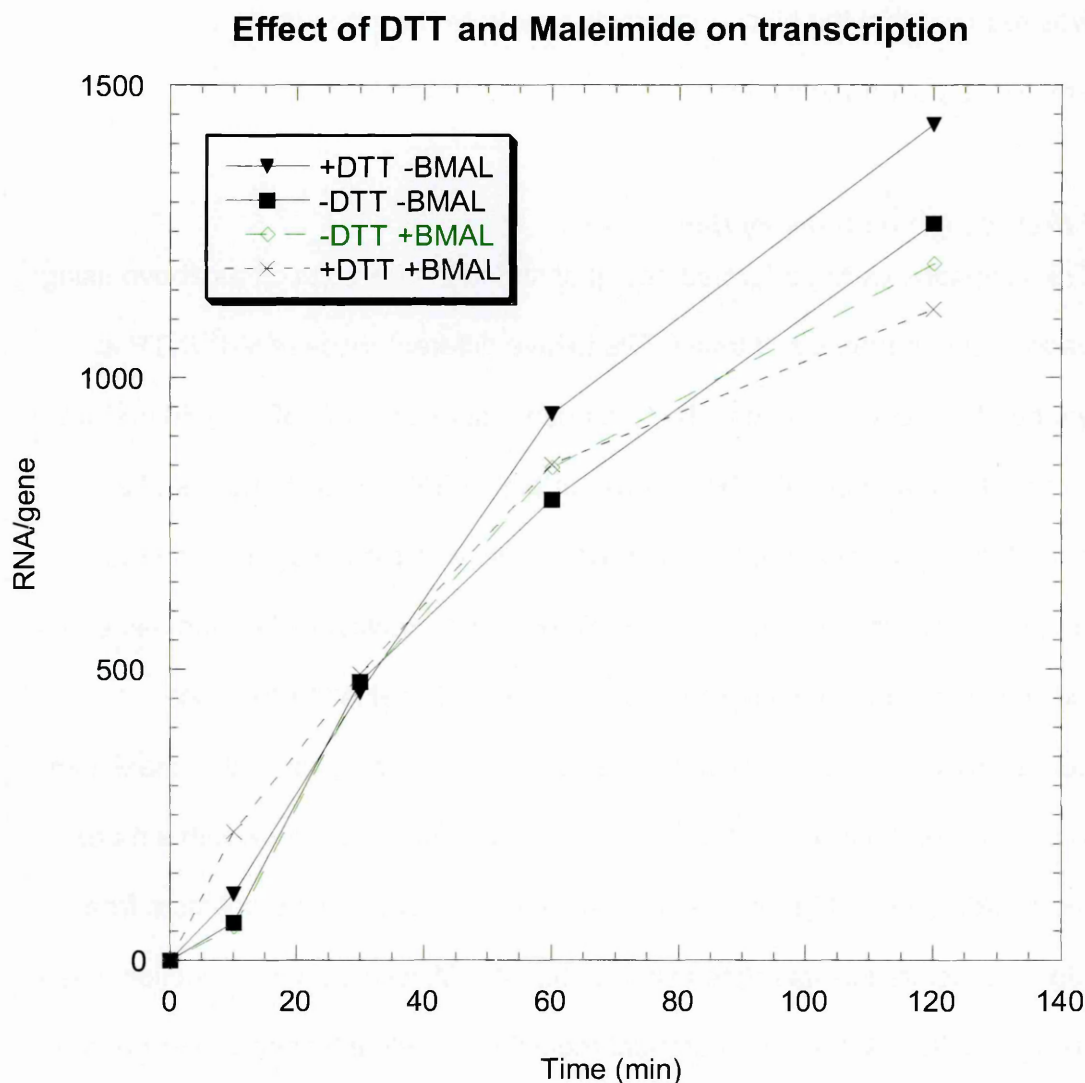


Figure 10 Transcription in the presence of DTT and Maleimide

Transcription levels were compared in the presence or absence of 50 mM DTT and 50 μ M biotin-maleimide.

As expected, the sample with added DTT and no added maleimide has the highest transcription levels, and the samples with maleimide have the lowest transcription levels. The differences are not large though, with only an ~30% difference between the lowest and highest treatments. However, after further enrichment experiments, it was determined that higher amounts of biotin-maleimide (~200-500 μ M) would be required for the catalytic selections, and at these levels there was no detectable transcription (data not shown). Therefore it

was required that the biotin-maleimide substrate must be added to the emulsions post-transcriptionally.

Effect of Mg^{2+} on transcription

Transcriptions were performed and quantitated in bulk solution as above using various concentrations of total NTPs to give different ratios of Mg^{2+} /NTP at various Mg^{2+} concentrations. Mg^{2+} concentrations of 5, 16, 30 and 50 mM were used with an amount of NTPs to give an Mg^{2+} /NTP ratio of 1, 1.25, and 2.5 at each Mg^{2+} concentration. First, the RNA yield was plotted against the ratio (Figure 11A), with no correlation visible ($R^2=.02$). However when the same data were plotted such that the RNA yield was plotted against the *free* Mg^{2+} concentration (Figure 11B) one can see a definite correlation with a maximum yield at around 5 mM free Mg^{2+} . A further experiment was done with a more restricted range of Mg^{2+} concentrations (Figure 11C), and the optimum free Mg^{2+} concentration was determined to be ~4 mM, with a sharp reduction when using less than 4 mM, but a gradual reduction in yield if higher concentrations are used.

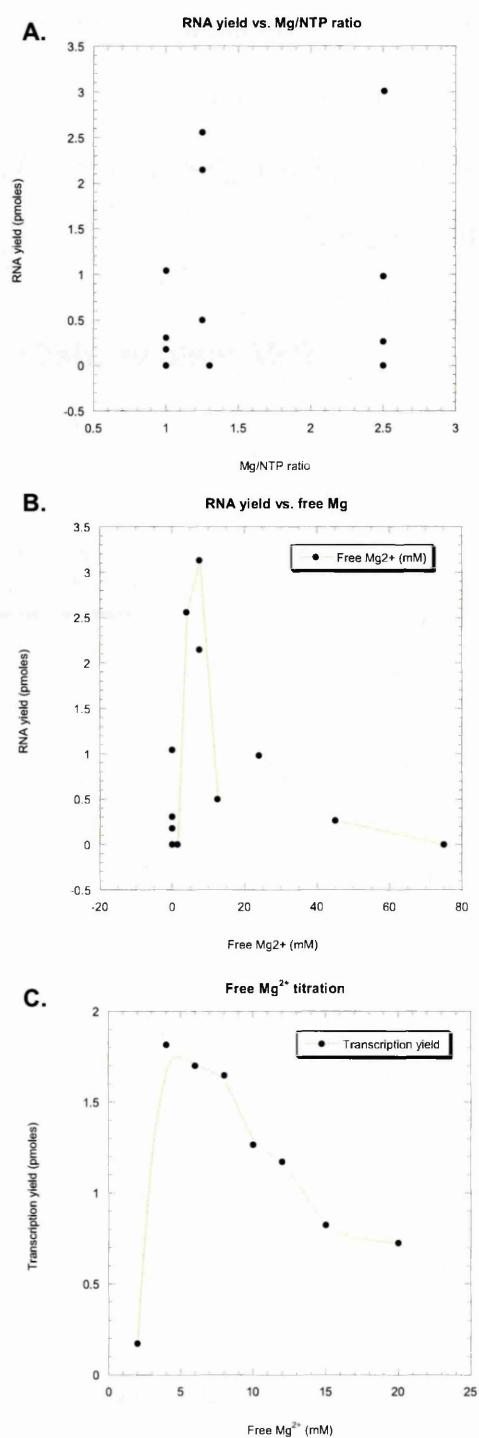


Figure 11 Effect of Mg²⁺ on transcription yield.

A. Yield vs. Ratio of Mg²⁺/NTP, B. Yield vs. free Mg²⁺. from the same experiment, C. Yield vs. Free Mg²⁺ at a smaller range.

Effect of NaCl concentration on transcription

Transcriptions were performed and quantitated in bulk solution as above using varying concentration of NaCl from 0-80 mM (Figure 12). There is a definite plateau at around 60 mM showing that higher salt concentrations would not be necessary for maximal transcription.

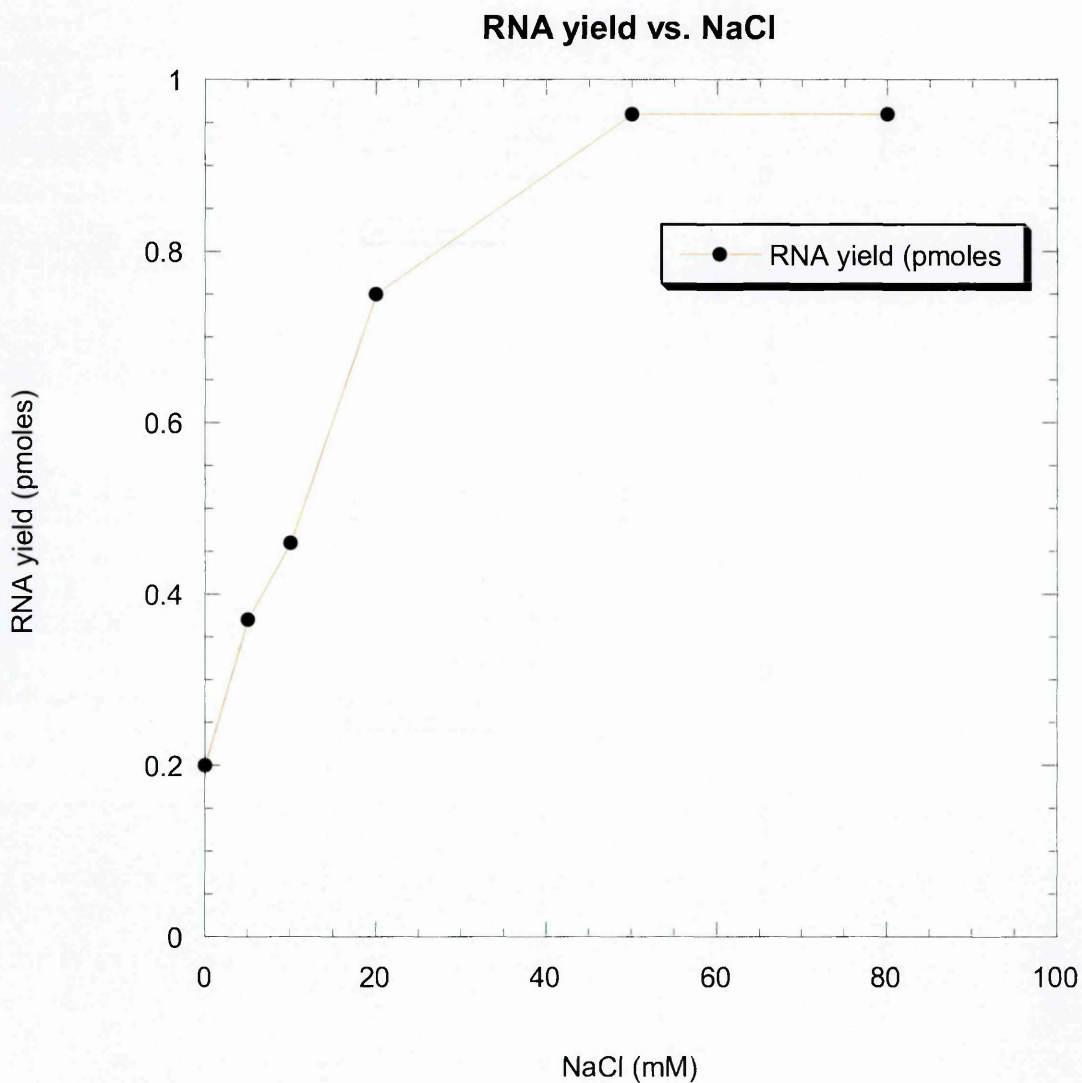


Figure 12 Effect of NaCl concentration of transcriptional yield
NaCl was varied between 0 and 80 mM.

Effect of emulsification conditions on transcription

Transcription yields were measured under different emulsification conditions. DNA at 0.33 nM (10^{10} genes in 50 μ l) was emulsified by 1) stirring in a 2 ml round-bottom Nunc cryotube with a Telesystem-15 stirrer, 2) Homogenization at 11 krpm with an UltraTurrax T-8 disperser, or 3) homogenization at 25.3 krpm. Transcription yields in bulk solution were compared to the various emulsification conditions by densitometry of PAGE gels as above (Figure 13). A decrease in RNA yield can be seen as the emulsification conditions become more intense, the rate of synthesis drops and in addition the total yield drops sooner for the more intense emulsifications. Emulsifying by stirring and homogenization at 11 krpm give similar yields until two hours, at which time the transcription rate in the homogenized sample begins to plateau. The transcription rate in the 25 krpm homogenization sample begins to plateau in less than one hour. This is likely due to either reaction components being consumed or inhibition from reaction products –such as pyrophosphate-, which would both be intensified in the smaller compartments of the homogenized samples. Note that the transcription in bulk solution shows a lower yield per gene than either of the two least intensely emulsified samples. This could be a concentration effect of the compartmentalization, since the DNA in bulk is only at 0.33 nM, but a single molecule within a compartment is likely be at a higher concentration (up to several nM depending on the size of the compartment). The increase the potential library size that can be screened due to both increased droplet number and lower polydispersity when homogenizing at 25 krpm, should counterbalance the small reduction (~50%) in transcriptional yields.

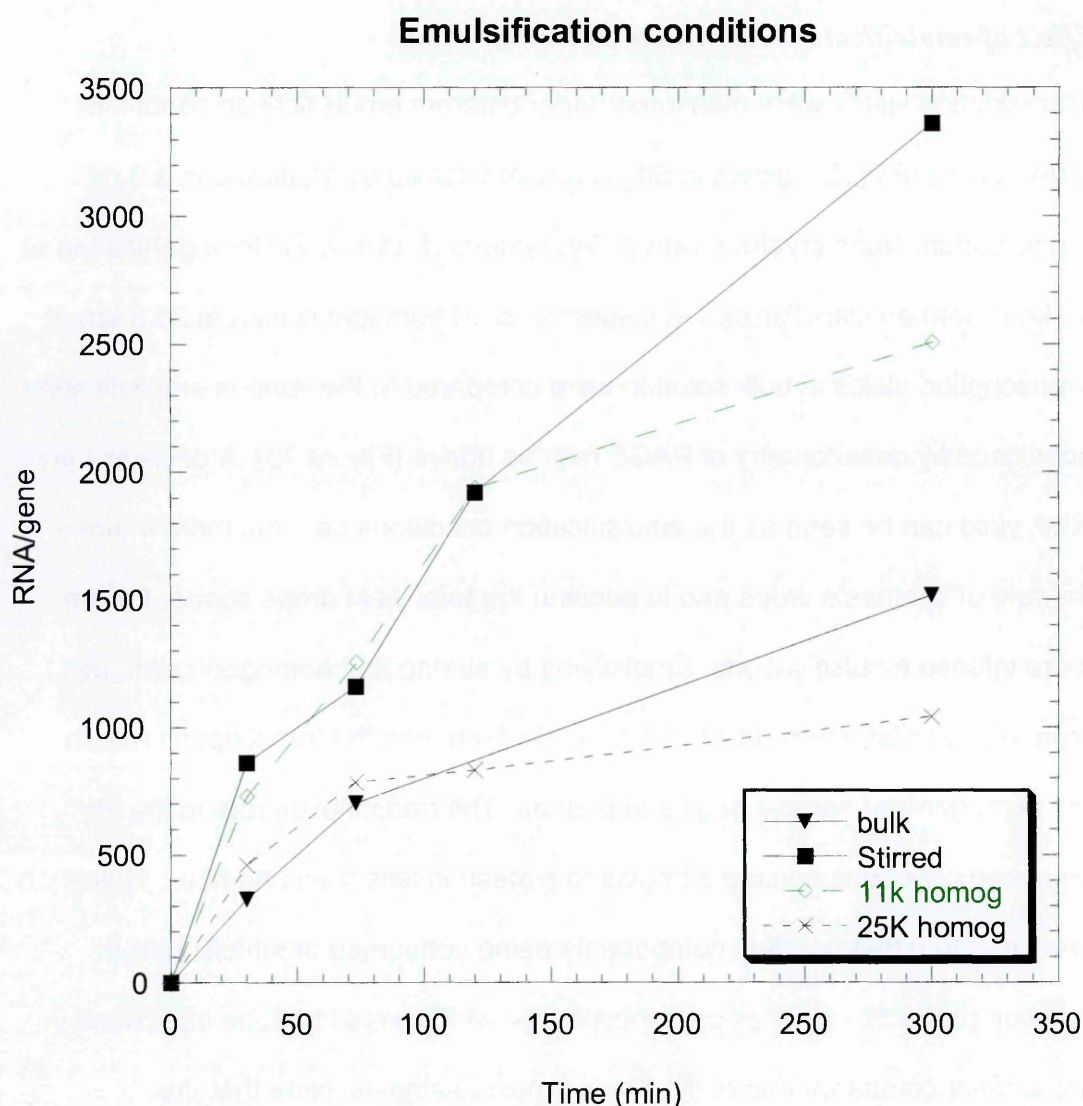


Figure 13 Transcription under different emulsification conditions

Transcription levels of different emulsification conditions are compared to bulk (unemulsified) transcription.

Optimisation of emulsified enrichment

Diels-alderase activity on DNA-coupled substrate

The anthracenylated oligo DA3-18 was annealed to its reverse complement (DA3-RC) to simulate the double-stranded DNA – anthracene conjugate that would be the Diels-Alderase substrate in the selection. The reaction was carried out with oligo concentrations of both 10 μ M and 1 μ M in the complete selection

transcription mix described in the Methods chapter including 200 μM biotin maleimide. The reactions were allowed to proceed for one hour both in the presence and absence of 7 μM purified 49mer ribozyme. An aliquot of each reaction was taken prior to the addition of the biotin maleimide substrate. The successful Diels-Alder addition of biotin-maleimide to the anthracene bound to the DA3-18 oligo is easily identified by a gel retardation of the adduct-coupled oligo relative to the substrate-coupled oligo (Figure 14). The amount of crude reaction loaded was adjusted so that the amount of substrate oligo on the gel was the same in each sample. approximately the same fraction of the substrate oligo (~10%) appears to have reacted for both substrate concentrations, which is what would be expected since the 10 μM is still at least an order of magnitude below the K_M for the anthracene substrate. No adduct-oligo is visible in the reactions without added ribozyme at either concentration, but in a subsequent experiment using 100 μM of anthracene-oligo, the adduct band was faintly visible after reacting for one hour in the absence of the ribozyme (data not shown). This formally demonstrates that the dsDNA-coupled anthracene is indeed a substrate for the Diels-alderase ribozyme.

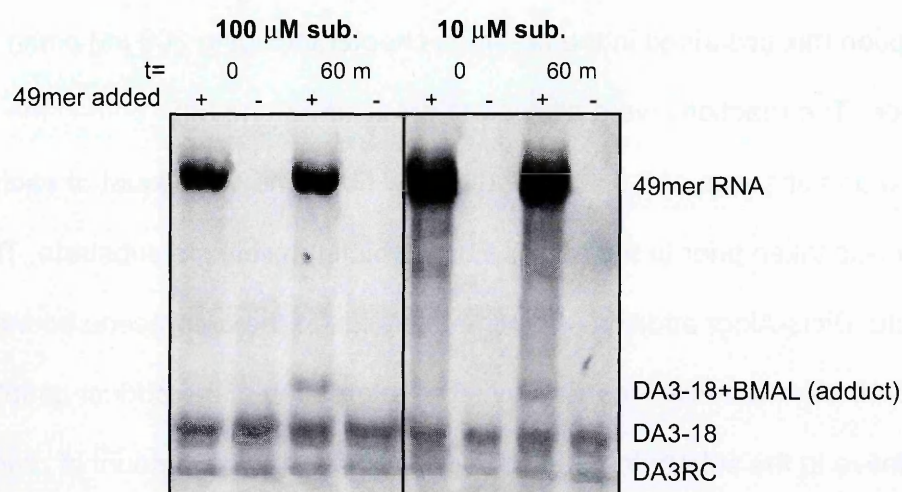


Figure 14 Ribozyme activity on double-stranded DNA.

The 49mer ribozyme was allowed to react with biotin-maleimide and anthracene-labeled double-stranded DNA (anthracenylated oligo DA3-18 annealed to its reverse complement DA3RC). The Diels-alder addition of biotin-maleimide to the DA3-18 oligo can be seen by the adduct's lower mobility on the gel. The 10 μM sample has been diluted 10-fold less than the 100 μM sample to normalize amount of substrate/product oligonucleotide loaded on the gel.

Effect of Mg^{2+} buffer and temperature on enrichment

A simplified selection was performed in bulk solution where anthracene-labelled genes encoding active 49mer Diels-Alderase genes (DA) were incubated with 7 μM purified 49mer ribozyme, and inactive deleted genes (ΔDA) were incubated in the same conditions without added ribozyme in a separate tube. After incubation the reactions were mixed in a 1:1 ratio and bound to streptavidin-coated dynabeads, washed, and amplified by PCR as described in the methods section. The PCR products (140 and 100 bp for the DA and ΔDA genes respectively) were resolved on an agarose gel, and the relative intensities of the bands quantified. An aliquot was also taken just prior to bead binding, and PCR amplified in the same way as the washed beads. The ratio of the DA: ΔDA bands before and after bead binding were compared, and the "fold enrichment" is the ratio of the post- and pre- bead binding intensities. Due to the nature of PCR this is not a quantitative measure of the exact level of enrichment, but it

does give a qualitative illustration of whether there is enrichment, and the relative magnitude of enrichment. See Figure 15 for an illustration of a gel using this method of detecting enrichment.

There is very little, if any, difference between the enrichment Tris-HCl pH7.4 and HEPES-KOH pH 7.0 as buffering compounds (Figure 16A). A similar observation was made by direct spectrophotometric monitoring of the reaction rate (see kinetic characterization methods), but in this experiment there was no observable difference in kinetic rate at all (data not shown). Therefore HEPES was chosen for future experiments due to a significant (2-fold) increase in transcriptional yield.

The effect of Mg^{2+} concentration is more dramatic (Figure 16A) with very little enrichment observed at the transcriptionally optimal Mg^{2+} concentration of 4 mM even with a relatively high (7 μ M) concentration of purified RNA as the catalyst. At 60 mM Mg^{2+} the enrichment is much greater due to the increased activity of the ribozyme. RNA requires Mg^{2+} to neutralize the negative charge of the sugar-phosphate backbone in order to fold into complex 3-dimensional shapes. In the case of the Diels-Alderase ribozyme, Mg^{2+} is not required for chemical catalysis, but it has been shown that high concentrations (~60 mM) are required for maximum activity, presumably due to folding requirements (3) (and personal observations).

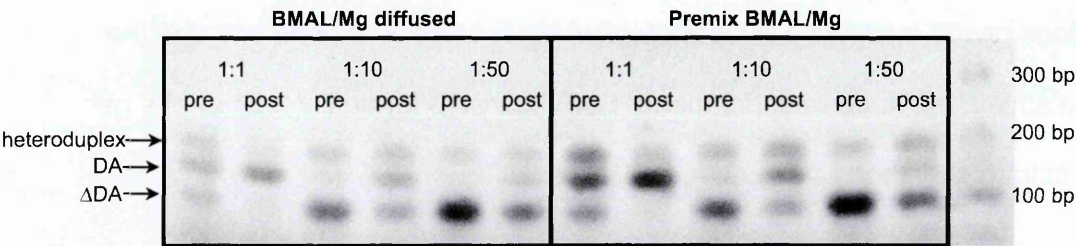


Figure 15 Mock selection of 49mer ribozyme.
Enrichment from before bead binding (pre) to after bead binding (post) is measured at different starting ratios of DA:ΔDA genes. Also, the effect of adding the biotin-maleimide substrate and Mg^{2+} after emulsification is compared to premixing the biotin-maleimide substrate and Mg^{2+} before emulsifying

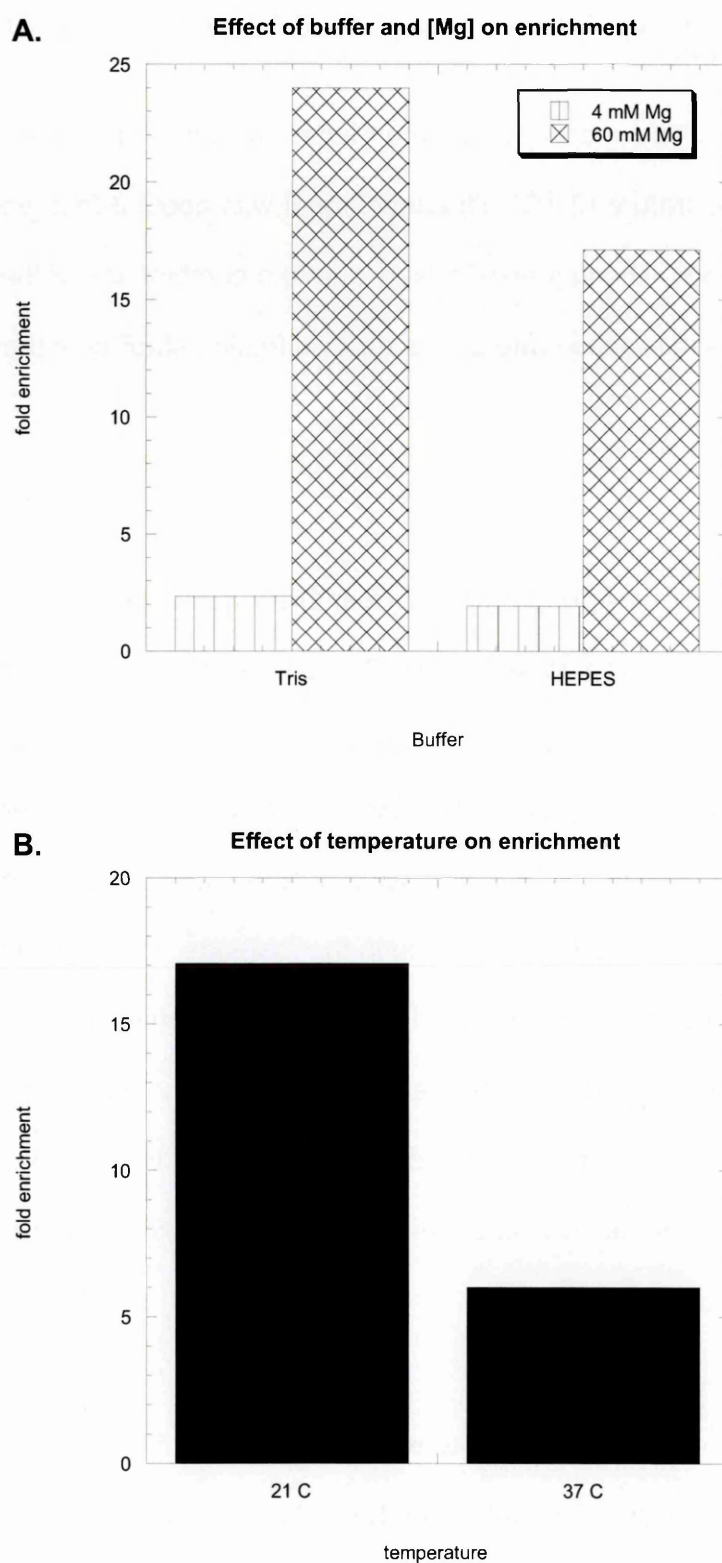


Figure 16 Effect of buffer Mg^{2+} and temperature on enrichment.

A. Enrichment factor is compared for different buffering compounds and different Mg^{2+} concentrations. **B.** Enrichment factor is compared for different temperatures. The enrichment

factors, based on differences in band intensities on a gel (Figure 15) are not meant to be fully quantitative (see text).

Transcription is optimal at 37°C, but the 49mer has always been assayed at or near room temperature (21°C). The enrichment was about 3-fold greater at 21°C (Figure 16B). The increase is likely due to a combination of the reduction in the uncatalysed reaction rate and enhanced folding stability of the 49mer at 21°C.

Mock selection

A “mock selection” was performed where anthracene-labelled DA genes and Δ DA genes were emulsified separately. The DA genes were emulsified with purified 49mer ribozyme (7 μ M) to simulate the transcription of active ribozymes in compartments encoding the 49mer. After emulsification the two emulsions were mixed in 1:1, 1:10, and 1:50 ratios (by weight), and the selection procedure was performed as described in the methods. The experiment was performed in parallel in one case adding 200 μ M biotin-maleimide and 60 mM Mg^{2+} to the reaction before emulsification, and in the other case adding these components after emulsification as described in the selection methods. In both sets of experiments the enrichment is approximately 10-fold (Figure 15).

Model selections

SELEX has previously been used to select a library of 157mer RNAs with the central 120 nucleotides randomized for Diels-Alderase ribozymes that catalyse the [4+2] cycloaddition of maleimide and anthracene (3). All of the ribozymes selected contained the same conserved ‘internal loop’ motif that was thought to

form the active site. This information was used to rationally design a minimized 49mer ribozyme that catalyses multiple-turnover Diels-Alder cycloadditions (85).

We have used this 49mer ribozyme to catalyse the selection of ribozymes by IVC (Figure 19). To allow amplification of ribozyme genes while retaining the possibility of randomizing the entire internal loop we created a gene that encodes a 69mer ribozyme, which is identical to the 49mer except for a 20-nucleotide extension at the 3' end (Figure 5). The kinetics of this ribozyme were nearly identical to that of the 49mer (data not shown). We also created a mutant, termed Δ 69mer, in which there is a 40 bp deletion encompassing the T7 promoter and the first 6 bp of the ribozyme gene (Figure 5).

Anthracene-labelled PCR products of the 69mer and Δ 69mer were made by amplifying the corresponding plasmids with an anthracene-labelled primer. The labelled PCR products were added to an *in vitro* transcription reaction mix in various ratios of 69mer to Δ 69mer genes at DNA concentrations of 3 pM – 3 nM (10^8 – 10^{11} total genes), emulsified, and put through the selection procedure described in Figure 4.

Selections were performed under both single- and multiple-turnover conditions. Selective pressure for multiple turnover can be achieved by adding free anthracene-HEG substrate to the emulsion compartments to a concentration of 100 μ M (a 10^4 -fold excess over the \sim 9 nM DNA tethered AHEG). Based on the second-order rate constant for the uncatalysed reaction (k_{uncat}) of 3 M⁻¹min⁻¹ (85), using 100 μ M untethered anthracene-HEG (a 10^4 -fold excess over the

DNA-tethered HEG) and 200 μM biotin-maleimide the expected rate of the uncatalysed reaction is 60 nM/min, and 540 nM of product – an amount equal to the RNA concentration – would be formed in nine minutes. Therefore, using a 9 min reaction time, genes for ribozymes that can perform a single turnover should be enriched 2-fold compared to genes that do not encode active ribozymes. In 9 min less than 1% of the substrate molecules should be turned over by the uncatalysed reaction. Hence, if a ribozyme can turn over multiple times, then the enrichment will be that much better, with a theoretical upper limit to the enrichment factor of ~ 100 -fold.

Due to different requirements of the transcription and catalytic reactions, and also to the relatively high rate of the uncatalysed reaction between the maleimide and anthracene substrates, it was necessary to alter some conditions within the emulsions after the emulsions were already formed. We have previously shown that it is possible to deliver a substrate and change pH in the compartments of a pre-formed water-in-oil emulsion (47). Furthermore, it is possible to deliver various solutes, including metal ions, into pre-formed emulsion droplets via a nano-droplet delivery system (49). The addition of MgCl_2 dissolved in 50% ethanol to the oil phase brought the free Mg^{2+} concentration within the compartments from a transcriptionally optimal 4 mM to a higher level necessary for catalytic activity (3). The biotin-maleimide substrate was also added after transcription for two reasons. First, transcription is inefficient in the presence of maleimide, probably due to the maleimide reacting with the active site cysteine in T7 RNA polymerase. Second, this prevents the uncatalysed reaction generating a significant amount of product before an adequate amount

of ribozyme has been transcribed, increasing the dynamic range of the selection. When either biotin-maleimide or Mg^{2+} are not added, no enrichment is observed (data not shown).

DNA was amplified by PCR pre- and post-selection (Figure 17). A starting ratio of 1:5 69mer: Δ 69mer genes was enriched to approximately a 1:1 ratio after selection (a five-fold enrichment) under both single- and multiple-turnover conditions (data shown for single-turnover only). No enrichment was observed when the reactions were not emulsified, demonstrating that, as expected for in *trans* catalysis, compartmentalization of the genes in an emulsion is essential for enrichment. This relatively low enrichment is entirely due to the relatively high rate of the uncatalyzed reaction (k_{uncat} of $3\text{ M}^{-1}\text{min}^{-1}$ 85)) (see discussion).

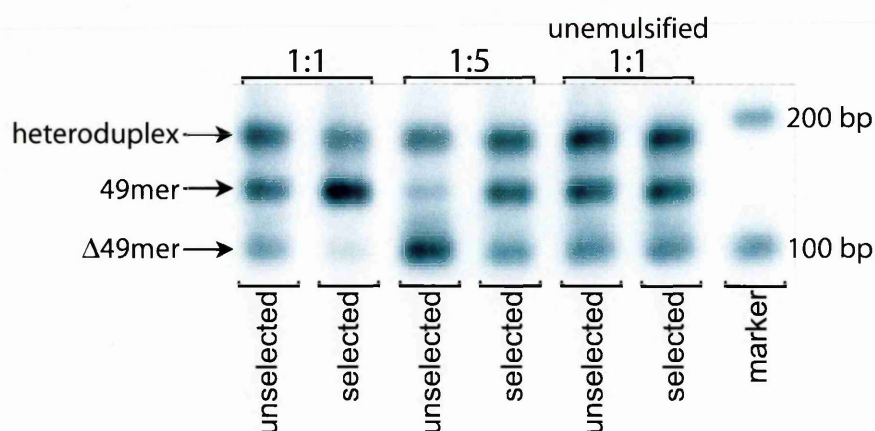


Figure 17 Model selection of 69mer ribozyme by IVC.

Genes from before and after selection were PCR amplified and analyzed by agarose gel electrophoresis. Selections were performed starting from 1:1 or 1:5 ratios of 69mer: Δ 69mer genes either emulsified or unemulsified.

Chapter 4

Selection of libraries

69mer libraries

In order to probe the extent to which the wt49mer represents the optimum catalytic solution for the trans reaction, several libraries were created where the predicted active site in the internal loop region was randomized. Four libraries were created based on the wt49mer sequence and predictions of its structure and active site (3,99), noting that several nucleotide positions within the putative active site were conserved based on single-mutation studies (99). Details of the libraries are given in Figure 18. Library 1 was chosen to randomize the entire internal loop and adjacent bases in the helices. In library 2 only the non-conserved positions (99) of the internal loop and adjacent helices were randomized. In library 3 the non-conserved positions of the internal loop only were randomized. In library 4 only the conserved positions were randomized.

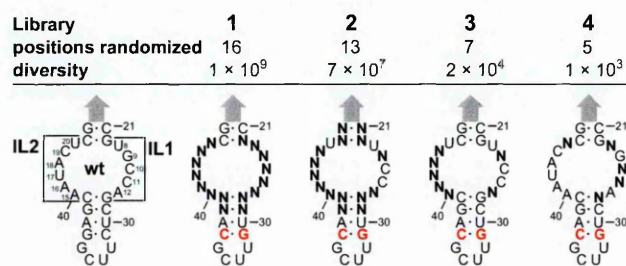


Figure 18 Libraries of the IL region of the 69mer ribozyme.

Randomized positions are written as "N". numbering outside of the IL region is consecutive starting at the 5' end. Within the IL region, numbering is as (3).

In order to randomize the internal loop region and still have space at the 3' end of the ribozyme gene for a constant primer binding site to amplify genes between selection rounds, a 20 bp extension was added to the 3' end of the gene using PCR to create the wt69mer gene (Figure 5e, g). Based on the predicted structure (99), the extension itself was expected to have a negligible effect on the catalytic activity of the ribozyme. In fact, the sequence of the extension was taken from one of the original SELEX clones that showed high activity (3).

To avoid contamination from wild-type sequences, the pIVEX plasmid template used for PCR amplification of the libraries contained only the T7 promoter with no ribozyme gene. Libraries were created by encoding the entire randomized ribozyme gene on a single PCR primer that annealed to pIVEX at the T7 promoter. This primer was paired with one annealing upstream of the T7 promoter in pIVEX to PCR amplify the selection construct.

A compensatory mutation was made in all of the libraries that changed the C31/G36 base pair to G31/C36 in order to allow identification of non-library-derived contamination (highlighted red in Figure 18).

A single anthracene molecule was attached to the distal end of each gene using an anthracene-labelled PCR primer. 10^{11} genes were selected as Figure 4 under both single- and multiple-turnover conditions.

All four libraries were first put through the single-turnover selection regime. After the fifth round of selection the pool from the smallest library, library 4, showed activity similar to that of the wt69mer, and so selection was stopped for this library at this point. The selected pool was ligated into a pIVEX, transformed into *E. coli*, and individual clones were screened for activity. Twenty-nine of the most active clones were sequenced. All of the sequenced clones had the C/G-G/C mutation confirming that they were not produced by external contamination. However, the sequence of the internal loop was the same as wild-type in every active clone. Besides a few rare spurious mutations in the terminal loop regions of the helices, which wouldn't be predicted to have a functional effect, the only area of variation in the selected clones was at the last base pair of helix III. At position twenty-eight (see Figure 18) 13/29 (45%) of the clones had the wild-type G; 11 (38%) had U; 2 (7%) had C, and none had A. These results are consistent with single-mutation studies of this position (99) for the cis reaction, which showed that any nucleotide could be substituted at this position with little effect on catalytic activity.

Selection for libraries 1-3 was continued after round five, now under multiple turnover conditions. After the ninth round of multiple-turnover selection, individual clones were cloned and screened for activity as above, and several from each library were sequenced. As with library 4 above, all of the sequenced

clones had the C/G-G/C mutation confirming that they were not produced by external contamination.

The three fastest and three of the inactive clones were sequenced from library 1. All of the fastest clones had the wild-type internal loop sequence, and all of the inactive clones had several mutations in this region.

All eight of the active clones sequenced from library 2 had the wild-type internal loop sequence. The only variation from the wild-sequence was the G28U mutation that was also seen in the single-turnover selections from library 4. A single inactive clone sequenced from this library had several mutations in the randomized region of the internal loop

Similarly, all nine of the active clones from library 3 had the wild-type sequence in the internal loop. There was one G28A, and two G28U mutations, once again showing the sequence variation allowed at this position.

157mer library

cDNA recovered after the 5th round of SELEX (3), which should already be enriched for active Diels-Alderase ribozymes, was amplified by PCR to generate DNA fragments in which transcription is driven off a proximal T7 promoter. A single anthracene molecule was attached to the distal end of each gene using an anthracene-labelled PCR primer. 10^{11} genes were selected as Figure 4 under both single- and multiple-turnover conditions.

Diels-Alderase activity of the library increased more quickly under single-turnover than under multiple-turnover conditions (Figure 19a). By round six of the single-turnover selections activity of the library was equal to that of the 69mer, while the activity of the library selected under multiple-turnover conditions had reached approximately half the activity of the 69mer by round nine. 53 individual clones with the highest activity from round six of the single-turnover selections and 11 clones with the highest activity round nine of the multiple-turnover selections were sequenced.

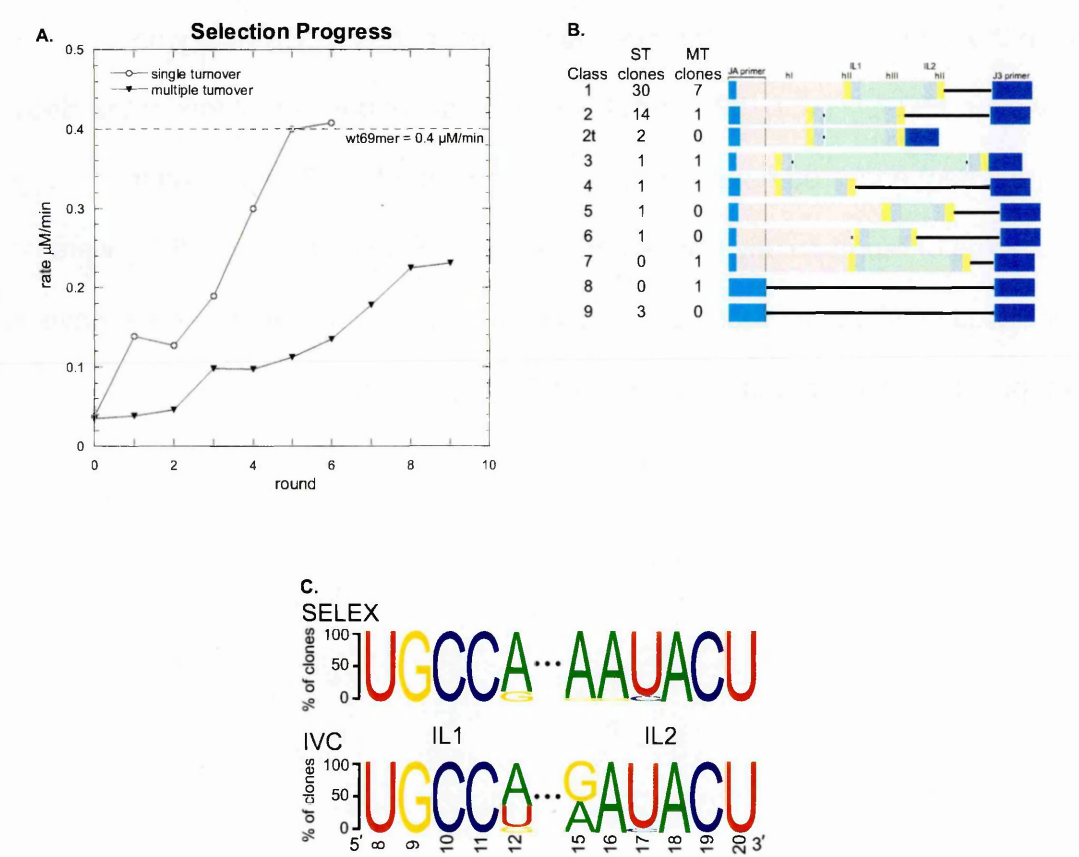


Figure 19 Selection of 157mer libraries by IVC.
Panel A. Selection progress under single- and multiple-turnover conditions. 5 nM DNA, amplified from each round of selection was transcribed in a 10 μl transcription reaction and the whole reaction assayed for Diels-Alderase activity with 100 μM AHEG and 500 μM biotin-maleimide by following the reduction in absorbance at 365 nm. Rates are plotted with the background rate due to the uncatalysed reaction subtracted. The dotted line indicates the rate of

the 69mer ribozyme (0.4 μ M/min). **Panel B.** Schematic showing the different classes of ribozymes isolated from the selections. The primer binding sites and, for the ribozymes possessing the same structural motif as the SELEX ribozymes (3), helices I, II and III (hI, hII and hIII) and the two halves of the internal loop (IL1 and IL2) of the ribozyme are labelled and shown as coloured bars. Regions with no homology to the SELEX ribozymes are shown with a black line. **Panel C.** Comparison of the internal loop sequence of ribozymes selected by SELEX and IVC displayed in format similar to a sequence logo (100).

All ribozymes isolated from the SELEX library after 11 rounds of selection shared the same basic secondary structure motifs (3,99). That is, an internal loop (IL) flanked by two helical regions (hII, and hIII) with a helical region (hI) extending from near the 5' terminus up to helix hII (Figure 20). Using IVC nine distinct sequence classes were identified (Figure 19b). Clones from Classes 1-7 have the same secondary structure motif found in the SELEX ribozymes. However, MFOLD (101,102) predictions of the secondary structures of the class 8 and 9 sequences do not appear to share this structure. The secondary structures of representative clones from class 1 (F12) and class 9 (E03) were confirmed by structural probing, indicating that class 9 ribozymes indeed have a completely novel fold (Figure 20; Figure 21, Figure 22).

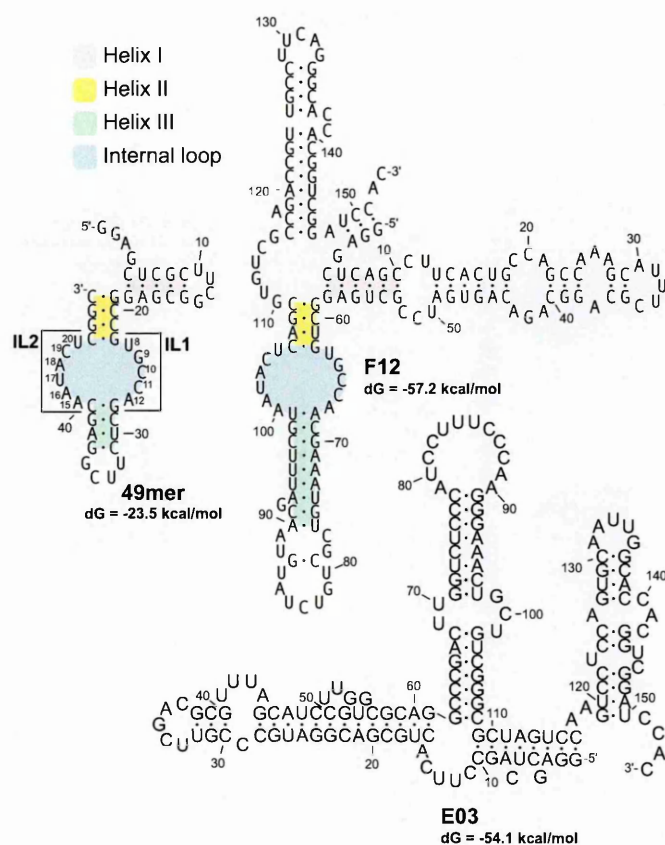


Figure 20 Secondary structure of ribozymes.

Lowest energy secondary structures as predicted by MFOLD (101,102) of the 49mer ribozyme (3), and two ribozymes from the IVC selections: F12 (class 1), which has the same structural motif and internal loop sequence as the 49mer ribozymes, and E03 (class 9), which has a novel secondary structure, not found in ribozymes from SELEX (3). The predicted secondary structures were confirmed by inline attack probing (89) and partial nuclease digestion using ribonucleases T1, T2, and V1 (103) (see Figure 21 and Figure 22). Helices I, II and III and the internal loop of the 49mer and F12 ribozymes are labelled. The numbering of the IL region in brackets is the same as in (3).

Clone F12

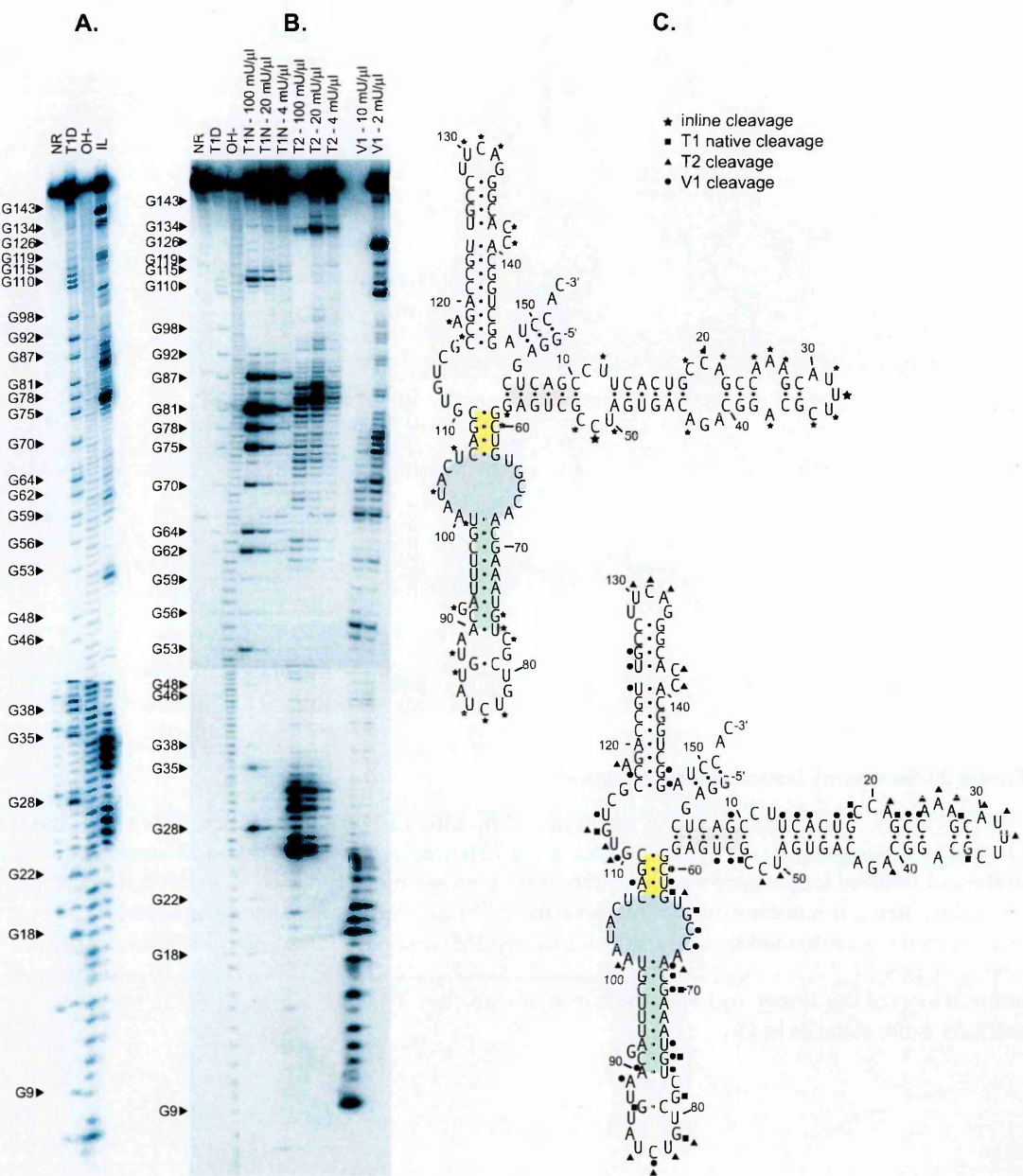


Figure 21 Secondary structure probing.
The secondary structure predicted using MFOLD (101,102) of ribozyme F12 (class 1) from the IVC selection was confirmed by both inline attack probing (89) and partial nuclease digestion using the ribonucleases T1, T2, and V1 (103), that cleave after ssG, any ss, and any ds residues respectively. *Panel A*, structure probing analysed on separate 8% denaturing polyacrylamide gels for inline attack and partial nuclease digestion methods. *Panel B*, cleavage sites on the ribozymes. Inline attack and nuclease cleavage points are marked with symbols on separate models for clarity.

Clone E03

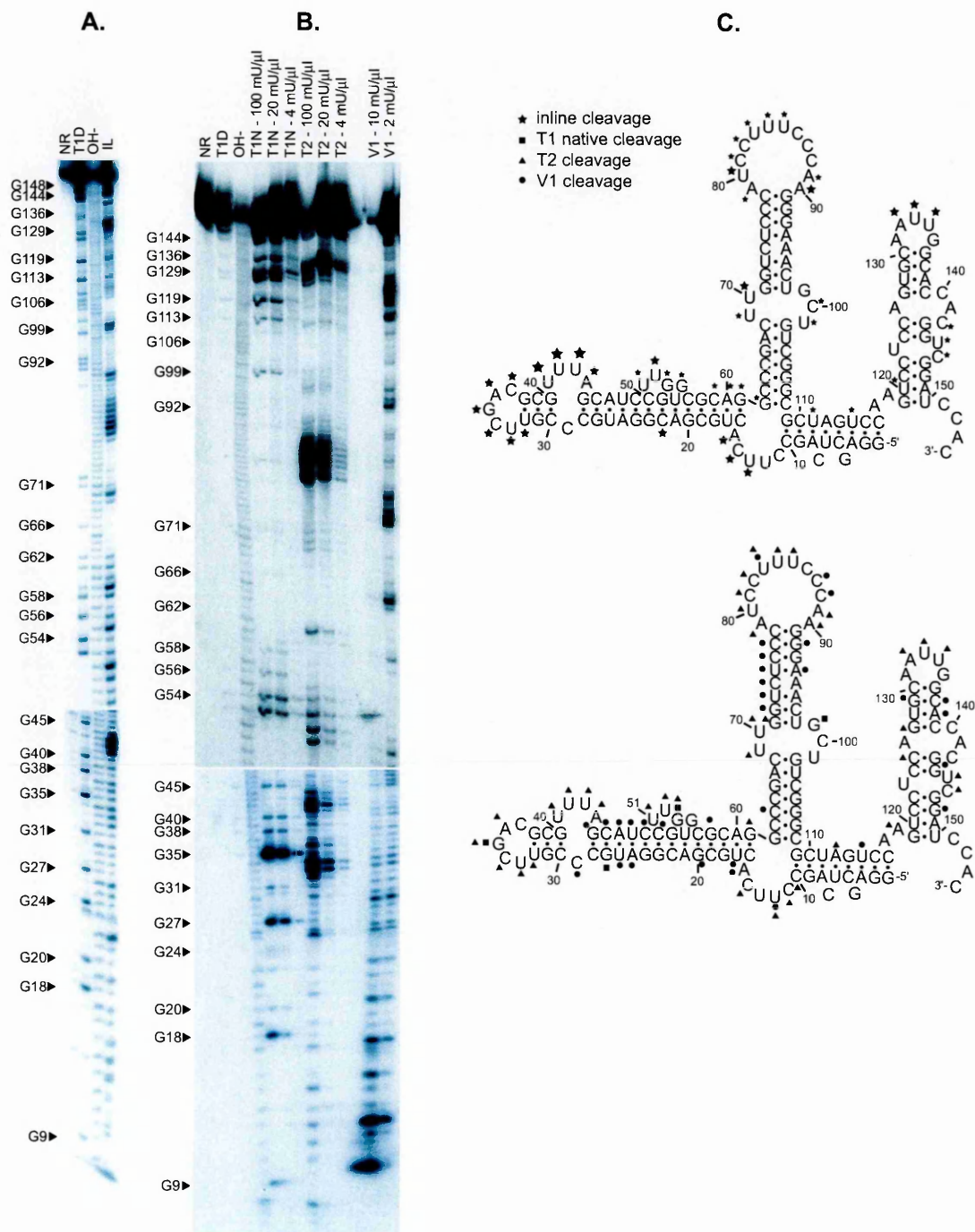


Figure 22 Secondary structure probing.

The secondary structure predicted using MFOLD (101,102) of ribozyme E03 (class 9) from the IVC selection was confirmed by both inline attack probing (89) and partial nuclease digestion using the ribonucleases T1, T2, and V1 (103), that cleave after ssG, any ss, and any ds residues respectively. *Panel A*, structure probing analysed on separate 8% denaturing polyacrylamide gels for inline attack and partial nuclease digestion methods. *Panel B*, cleavage sites on the

ribozymes. Inline attack and nuclease cleavage points are marked with symbols on separate models for clarity.

There are 37 sequences in class 1 (57% of the total), and although there are point mutations scattered throughout the sequences, the internal loop region – the location of the putative active site (3,99) – is mostly conserved. The nucleotides in the IL region are numbered according to (3), and can be seen in Figure 20. The first part of the internal loop (IL1, positions 8-12) is conserved and has the sequence UGCCA, the same as the wt49mer. There is some variation in the distal part of the internal loop (IL2, positions 15-20) with 10/37 having the wt49mer sequence AAUACU, and 27/37 GAUACU. Notably, all of 27 of these A15G mutants come from the single-turnover selections.

Class 2 contains 17 sequences (26% of the total). As above, there are point mutations scattered around the sequences, and in this class the internal loop region is entirely conserved. The sequence of IL1 is UGCCU, and IL2 has the wt AAUACU. Although it does share the same basic structural motif with the SELEX-selected mutants, class 2 is not identical to any of the SELEX ribozymes.

Class 3 contains two sequences and is identical to SELEX group 7. Class 4 contains 2 sequences, and is not identical to any of the SELEX groups. Class 7 Contains one sequence and is identical to SELEX group 5. The remaining classes all have one sequence each, and are not identical to any of the SELEX groups.

The data for all of the IL region sequences from all of the ribozymes selected by both SELEX (40 clones sequenced) and IVC (64 clones sequenced) are displayed in Figure 19c in a format analogous to a sequence logo (100) with the

height of a particular letter in a stack proportional to the frequency of its occurrence in the selected clones. From this it is possible to see the difference in the spectrum of sequences selected by IVC and SELEX. The U12 variant that is the most common in IVC is not seen at all in the SELEX clones. The G15 variant that is the most common in IVC is only present in one clone from the SELEX selections. The complete DNA sequences of all of the IVC-selected clones are shown in Figure 23 with the regions correlating to 49mer secondary structure elements coloured.

Class	JA primer	J3 primer	ST clones	MT clones
1.	GGAGCTCAGCCTTCACTGCGAGGCAAGCATTGCGAGGAGACAGTATCGCTGAGGCTGTGCGACGAAGTGTGCTGTCTATGTACGATTTCGTGATCTACGCTGTGCGCAGGCGGTTGCTTCAG	GGAGCTCAGCCTTCACTGCGAGGCAAGCATTGCGAGGCTGTGCGACGAAGTGTGCTGTCTATGTACGATTTCGTGATCTACGCTGTGCGCAGGCGGTTGCTTCAG	30	7
2.	GGAGCTCAGCCTTCACTGCGAGGCAAGCATTGCGAGGCTGTGCGACGAAGTGTGCTGTCTATGTACGATTTCGTGATCTACGCTGTGCGCAGGCGGTTGCTTCAG	GGAGCTCAGCCTTCACTGCGAGGCAAGCATTGCGAGGCTGTGCGACGAAGTGTGCTGTCTATGTACGATTTCGTGATCTACGCTGTGCGCAGGCGGTTGCTTCAG	14	1
2L	GGAGCTCAGCCTTCACTGCGAGGCAAGCATTGCGAGGCTGTGCGACGAAGTGTGCTGTCTATGTACGATTTCGTGATCTACGCTGTGCGCAGGCGGTTGCTTCAG	GGAGCTCAGCCTTCACTGCGAGGCAAGCATTGCGAGGCTGTGCGACGAAGTGTGCTGTCTATGTACGATTTCGTGATCTACGCTGTGCGCAGGCGGTTGCTTCAG	2	0
3.	GGAGCTCAGCCTTCACTGCGAGGCAAGCATTGCGAGGCTGTGCGACGAAGTGTGCTGTCTATGTACGATTTCGTGATCTACGCTGTGCGCAGGCGGTTGCTTCAG	GGAGCTCAGCCTTCACTGCGAGGCAAGCATTGCGAGGCTGTGCGACGAAGTGTGCTGTCTATGTACGATTTCGTGATCTACGCTGTGCGCAGGCGGTTGCTTCAG	1	1
4.	GGAGCTCAGCCTTCACTGCGAGGCAAGCATTGCGAGGCTGTGCGACGAAGTGTGCTGTCTATGTACGATTTCGTGATCTACGCTGTGCGCAGGCGGTTGCTTCAG	GGAGCTCAGCCTTCACTGCGAGGCAAGCATTGCGAGGCTGTGCGACGAAGTGTGCTGTCTATGTACGATTTCGTGATCTACGCTGTGCGCAGGCGGTTGCTTCAG	1	1
5.	GGAGCTCAGCCTTCACTGCGAGGCAAGCATTGCGAGGCTGTGCGACGAAGTGTGCTGTCTATGTACGATTTCGTGATCTACGCTGTGCGCAGGCGGTTGCTTCAG	GGAGCTCAGCCTTCACTGCGAGGCAAGCATTGCGAGGCTGTGCGACGAAGTGTGCTGTCTATGTACGATTTCGTGATCTACGCTGTGCGCAGGCGGTTGCTTCAG	1	0
6.	GGAGCTCAGCCTTCACTGCGAGGCAAGCATTGCGAGGCTGTGCGACGAAGTGTGCTGTCTATGTACGATTTCGTGATCTACGCTGTGCGCAGGCGGTTGCTTCAG	GGAGCTCAGCCTTCACTGCGAGGCAAGCATTGCGAGGCTGTGCGACGAAGTGTGCTGTCTATGTACGATTTCGTGATCTACGCTGTGCGCAGGCGGTTGCTTCAG	1	0
7.	GGAGCTCAGCCTTCACTGCGAGGCAAGCATTGCGAGGCTGTGCGACGAAGTGTGCTGTCTATGTACGATTTCGTGATCTACGCTGTGCGCAGGCGGTTGCTTCAG	GGAGCTCAGCCTTCACTGCGAGGCAAGCATTGCGAGGCTGTGCGACGAAGTGTGCTGTCTATGTACGATTTCGTGATCTACGCTGTGCGCAGGCGGTTGCTTCAG	0	1
8.	GGAGCTCAGCCTTCACTGCGAGGCAAGCATTGCGAGGCTGTGCGACGAAGTGTGCTGTCTATGTACGATTTCGTGATCTACGCTGTGCGCAGGCGGTTGCTTCAG	GGAGCTCAGCCTTCACTGCGAGGCAAGCATTGCGAGGCTGTGCGACGAAGTGTGCTGTCTATGTACGATTTCGTGATCTACGCTGTGCGCAGGCGGTTGCTTCAG	0	1
9.	GGAGCTCAGCCTTCACTGCGAGGCAAGCATTGCGAGGCTGTGCGACGAAGTGTGCTGTCTATGTACGATTTCGTGATCTACGCTGTGCGCAGGCGGTTGCTTCAG	GGAGCTCAGCCTTCACTGCGAGGCAAGCATTGCGAGGCTGTGCGACGAAGTGTGCTGTCTATGTACGATTTCGTGATCTACGCTGTGCGCAGGCGGTTGCTTCAG	3	0

Figure 23 Diels-Alderase sequences selected by IVC.
The consensus of each sequence class from the IVC selection are shown with regions homologous to secondary structure motifs of the SELEX ribozymes (3) highlighted with colored bars as in Figure 19b. Helices I, II and III are pink, yellow and green respectively. The internal loop is grey, and the primer binding sites not involved in secondary structure are blue. Un-highlighted nucleotides are not involved in secondary structure analogous to the SELEX motifs. The numbers of observed sequences under the single-turnover (ST) and multiple turnover (MT) selection regimes are also listed.

Characterization of selected clones

Representatives from several classes of ribozyme, including representative clones with the same basic structural motif as the SELEX ribozymes (85) [clones F12 (class 1 wt sequence) E12 (class 1 A15G mutant) and G11 (class2)] and the novel class 9 fold (clone E03) were subjected to kinetic characterization. The kinetics of the 49mer ribozyme (85) were also measured under the same conditions (Figure 24, Figure 25, Figure 26, and Table 3). The kinetic parameters of ribozyme F12, were very similar to the 49mer ribozyme, which has the same structural motif and internal loop sequence, despite its greater length (157 bases). The other two ribozymes with the same basic structural motif, but with internal loop sequences which were seen rarely (clone E12) or never (G11) in SELEX, were slightly less efficient, both in terms of k_{cat} and $k_{cat}/K_{m1(diene)}K_{m2(dienophile)}$. Clone E03, with the novel fold, had very similar kinetic parameters to clone F12 and the 49mer ribozyme.

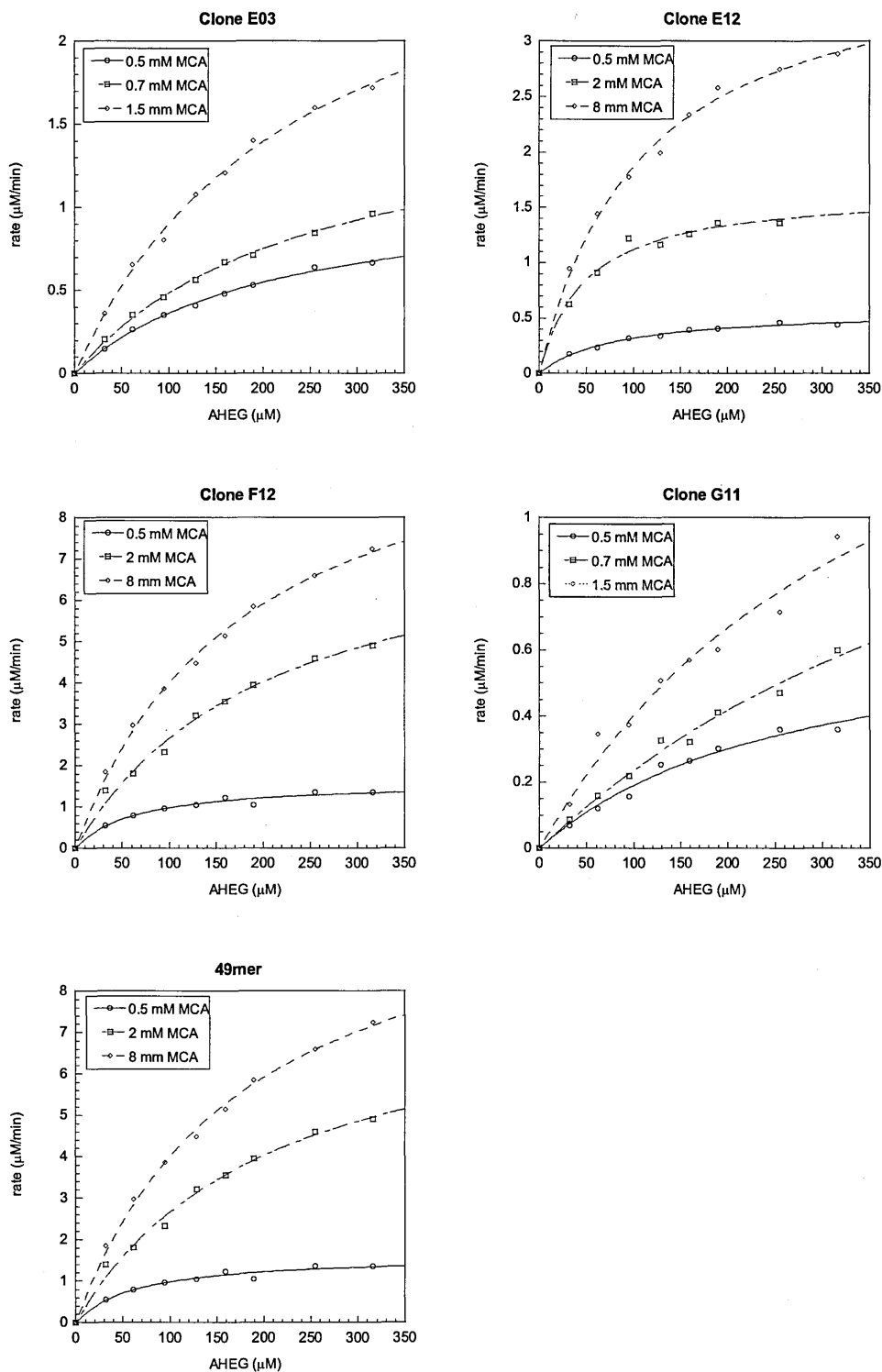


Figure 24 Kinetic plots of IVC-selected clones.

Plots show initial rate dependence on the diene concentration (AHEG) for several fixed concentrations of dienophile (MCA) (plots for only three of eight MCA concentrations are shown). Least-squares fits to the Michaelis-Menten equation are shown. Full kinetic details are given in Table 3.

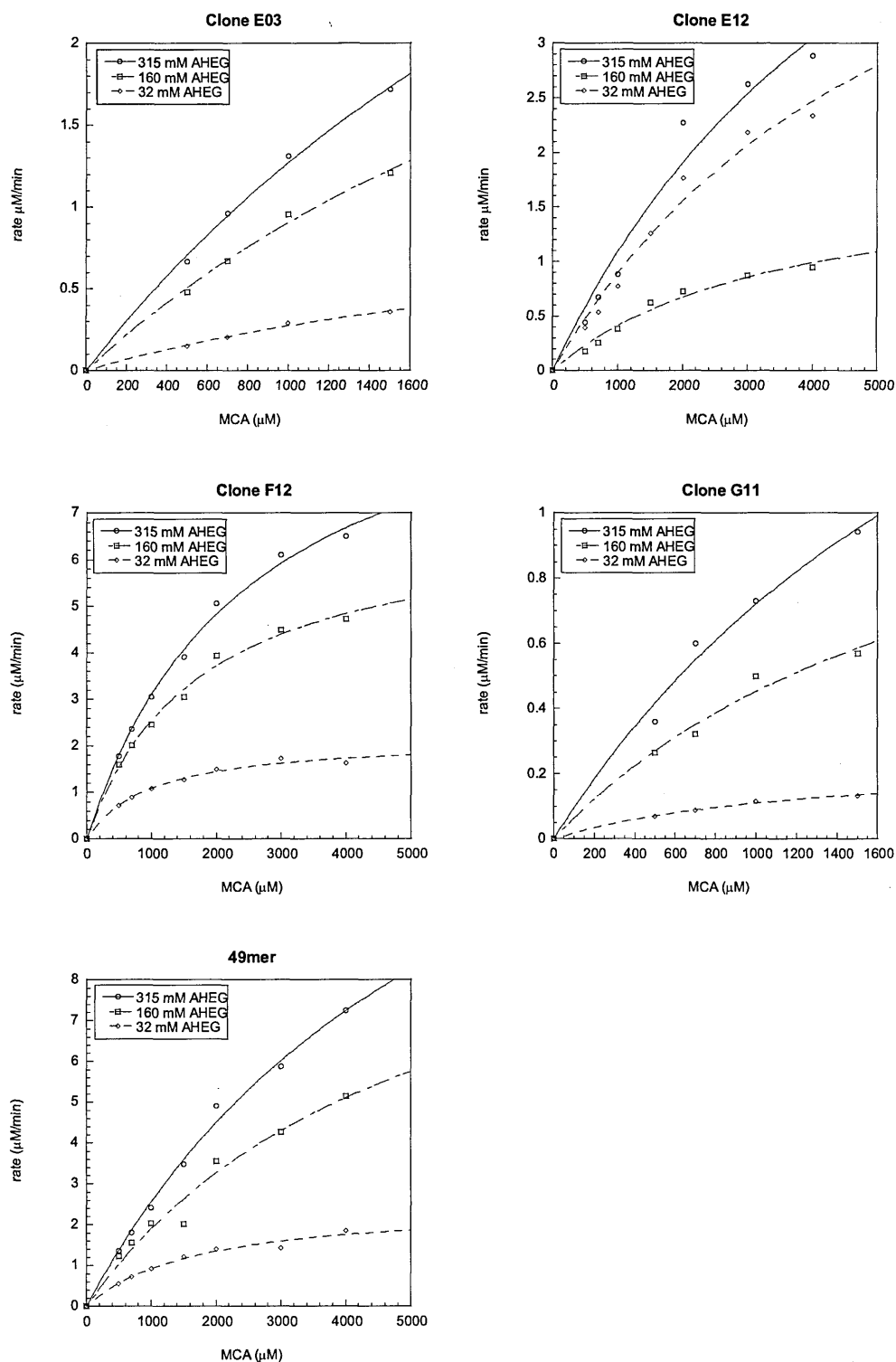


Figure 25 Kinetic plots of IVC-selected clones.

Plots show initial rate dependence on the dienophile concentration (MCA) for several fixed concentrations of diene (AHEG) (plots for only three of eight AHEG concentrations are shown). Least-squares fits to the Michaelis-Menten equation are shown. Full kinetic details are given in Table 3.

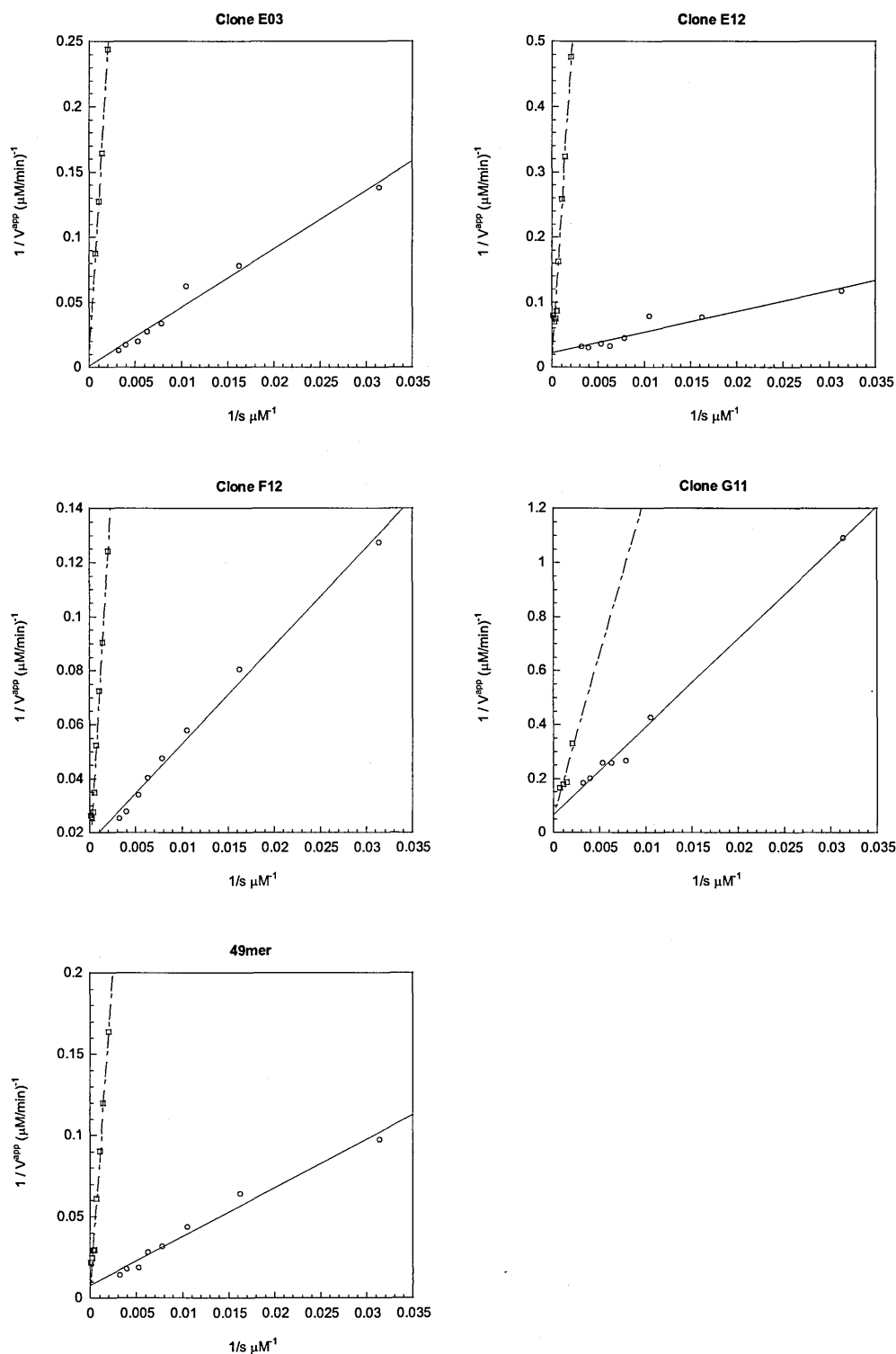


Figure 26 Secondary kinetic plots of IVC-selected clones. Plots of the apparent maximum rate (V^{app}) values for each substrate (s) (circles AHEG, squares MCA). The reciprocal plots are shown for clarity, but the V_{max} values for Table 3 in the main text were actually calculated by fitting the data to Equation 2.

Table 3 Kinetic parameters of Diels Alderase ribozymes.

Clone	Class	K_{m1} (diene) (μM)	K_{m2} (dienophile) (μM)	k_{cat} (s^{-1})	k_{cat}/K_{m1} ($\text{M}^{-1}\text{s}^{-1}$)	k_{cat}/K_{m2} ($\text{M}^{-1}\text{s}^{-1}$)	$k_{cat}/K_{m1}K_{m2}$ ($\text{M}^{-2}\text{s}^{-1}$)	K_i^\dagger (μM)
49mer*	1	200 \pm 35	5200 \pm 1300	0.33 \pm 0.06	1667	64.1	3.2 $\times 10^5$	11 (IC_{50}) [‡]
49mer	1	148 \pm 32	3960 \pm 623	0.32 \pm 0.11	2175	81.4	5.5 $\times 10^5$	31.4 \pm 1.8
F12	1	150 \pm 32	1650 \pm 197	0.13 \pm 0.04	878	79.9	5.3 $\times 10^5$	29.3 \pm 2.4
E12	1 [§]	86 \pm 10	7890 \pm 646	0.05 \pm 0.02	583	6.4	7.4 $\times 10^4$	57.3 \pm 2.9
G11	2	396 \pm 40	1310 \pm 65	0.01 \pm 0.003	25	7.6	1.9 $\times 10^4$	40.3 \pm 9.4
E03	8	213 \pm 7	6150 \pm 618	0.11 \pm 0.04	504	17.5	8.2 $\times 10^4$	68.0 \pm 2.7

*As measured in (104) (K_i) and (85) (all other parameters). [‡]Inhibition constant measured with the chemically synthesized Diels-Alder product racemate rac4 in all cases except [‡] where it was measured with the purified R,R-enantiomer of the reaction between 2 and N-pentylmaleimide (104). [§]Contains A15G change in internal loop sequence compared to 49mer and F12.

All the ribozymes were strongly product inhibited by the chemically synthesized

Diels-Alder product rac4 (Figure 8c) with the clones with sequences rarely or

never seen in the SELEX (clones E12, G11 and E03) having slightly higher K_i

values (Figure 27)

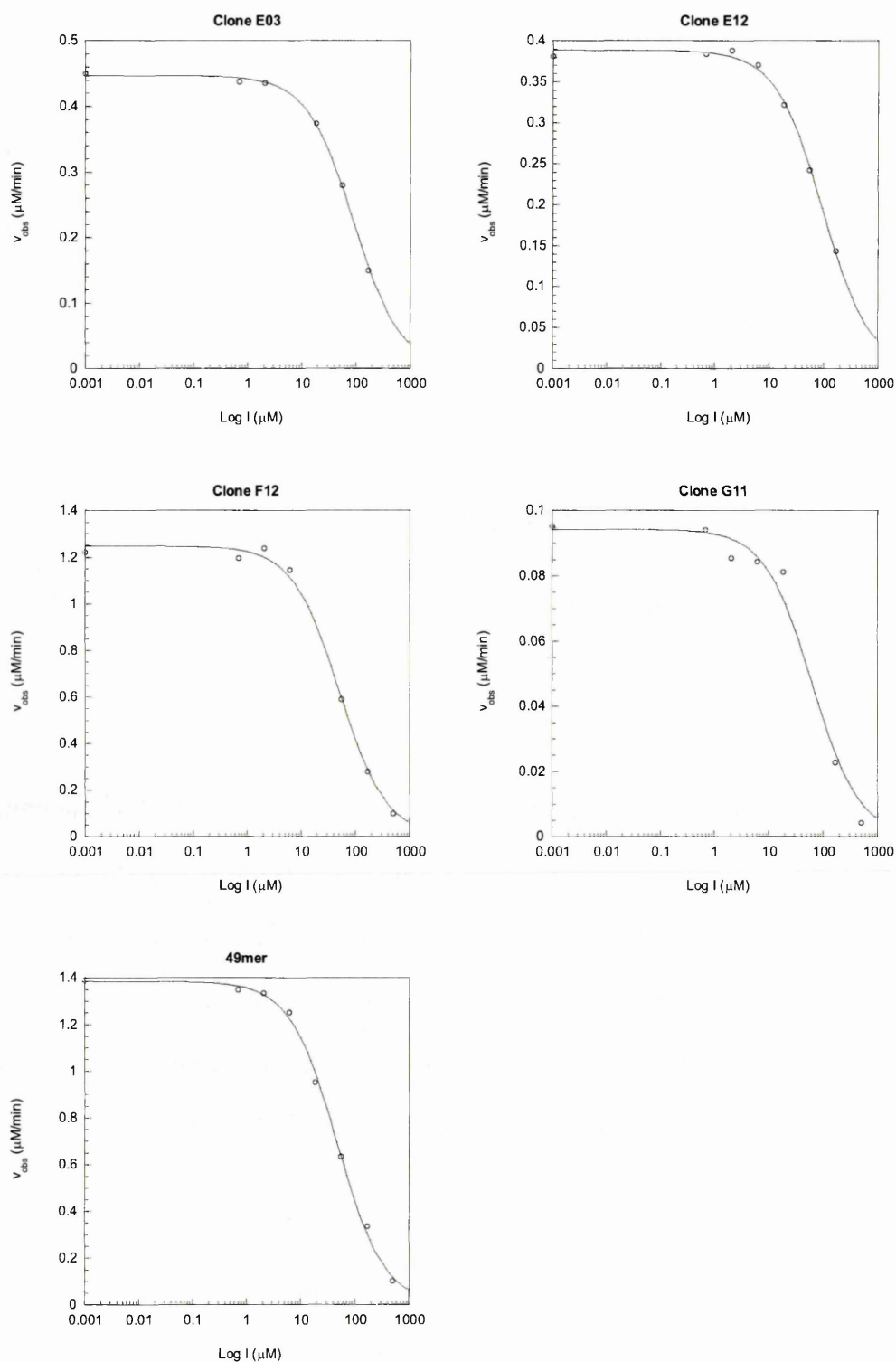


Figure 27 Product Inhibition plots of IVC-selected clones.

Plots show the reduction in initial rate for increasing concentrations of the chemically synthesized Diels-Alder product racemate rac4. Full kinetic details are given in Table 3.

Chapter 5

Discussion

To date, most ribozymes produced by *in vitro* selection have relied on the same SELEX-based selection strategy in which a substrate is physically attached to the RNA, or the RNA molecule is itself a substrate (58,69). Selection is for an intramolecular single-turnover reaction (in *cis* reaction), and active RNAs are enriched by selection for the formation of a product, which remains tethered to the RNA and allows either selective amplification or affinity purification of the RNA.

Here we describe, for the first time, the selection of ribozymes for true intermolecular catalysis (in *trans* reaction). This was achieved using *in vitro* compartmentalisation (IVC) of genes encoding ribozymes in aqueous droplets dispersed in a water-in-oil (w/o) emulsion as described in Figure 4.

These two approaches are very different. SELEX allows the selection of very large libraries of more than 10^{15} variants and can even select RNAs with very low rate accelerations ($k_{\text{cat}}/k_{\text{uncat}}$) (see(22)). However, selection is not truly for catalysis, since, by definition, a catalyst must be regenerated at the end of the catalytic cycle. There is selection pressure for product formation and rate acceleration, but not for product release or multiple turnover. Consequently, in many cases the selected nucleic acids have been found to accelerate the rate of transformation of a soluble, untethered substrate, and release the subsequent product, thus showing turnover. But the turnover shown by these ribozymes is typically quite limited (<20). However, multiple-turnover reactions have been observed in SELEX-selected ribozymes. This has been mainly with nucleic acid substrates where substrate recognition is driven by basepairing (for example, see 105), or only when nucleic acids are attached to the substrate (3). With other substrates, catalysis may not be observed when the substrate is detached from the RNA (for example, see 3,68). This is because the effective concentration of the substrate tethered to the RNA is very high and there is little or no selection for substrate binding.

IVC selects ribozymes that act upon substrates that are not tethered to the RNA itself. Hence, selection is for substrate recognition, rate acceleration and product formation simultaneously. Furthermore, selection for multiple turnover can be achieved by the inclusion of additional substrate to the compartments.

SELEX has previously been used to select both pyridyl-modified (68,69) and unmodified RNAs (3) which catalyse the Diels-Alder reaction. In the latter study,

ribozymes were selected that catalyse the Diels-Alder cycloaddition of anthracene (the diene) and maleimide (the dienophile). Here, we randomized the minimised 49mer ribozyme from that selection in an attempt to select for higher catalytic activity under intermolecular, multiple-turnover conditions. In addition, we used IVC to select exactly the same RNA library for the ability to catalyse the same reaction in order to directly compare the two methods.

We first optimised the IVC selection procedure using a ribozyme based on a minimised 49mer Diels-Alderase ribozyme which was rationally designed to incorporate the catalytic motifs identified in ribozymes selected using SELEX (3). This ribozyme catalyses the cycloaddition reaction between anthracene and maleimide free in solution (in *trans* reaction) with multiple turnover and high enantioselectivity (85). Selecting genes encoding the ribozyme from a background of a gene which does not (due to a 40 bp deletion encompassing the T7 promoter and the first 6 bp of the ribozyme gene) an ~5-fold enrichment was observed under both single and multiple-turnover conditions. This relatively low enrichment is entirely due to the relatively high rate of the uncatalysed reaction (k_{uncat} of $3.2 \text{ M}^{-1}\text{min}^{-1}$; (85)). Making the assumption that only a small amount of the substrates would be consumed during the reaction time, the expected enrichment factor will be equivalent to the ratio of initial rates of the catalysed and uncatalysed reaction (22), and can be expressed as:

Equation 6

$$\frac{V_{\text{cat}}}{V_{\text{uncat}}} = \frac{k_{\text{cat}}}{k_{\text{uncat}}} \cdot \frac{[E]_0}{K_{M1}K_{M2} + K_{M1}[S_2] + K_{M2}[S_1] + [S_1][S_2]} \quad (84)$$

Using the measured kinetic values for the wt49mer (Table 3), the theoretical enrichment with RNA concentration of 540 nM (60 transcripts/compartment), and diene (S1) and dienophile (S2) concentrations of 9 nM (one gene/compartment) and 200 μ M respectively would be 5.2-fold under single turnover conditions. Under multiple turnover condition, with 100 μ M diene (S1) (a 10^4 -fold excess of untethered substrate to the DNA-bound substrate) the theoretical enrichment would be 3.2-fold. These theoretical enrichments are very close to the observed enrichments. Six rounds of selection would still be expected to give a 1.6×10^4 -fold enrichment.

69mer selections

To probe in a combinatorial way the extent to which the SELEX-selected motif represented the optimal solution for catalysis under intermolecular, multiple turnover conditions we made several libraries where increasing fractions of the putative active site and the adjacent region were randomized with repertoire sizes of between $\sim 10^3$ and $\sim 10^9$. The libraries were initially selected for the *trans* reaction under single turnover conditions for five rounds with the goal of selecting catalysts that could turnover at least once. Then, the three larger libraries were selected for an additional nine rounds under multiple-turnover conditions.

The libraries were chosen based on information from single-mutation studies for the *cis* reaction of the 49mer ribozyme (99) that showed four positions where the base present in the wild-type could not be changed to any other without complete loss of activity, probably due to key structural interactions. Library 4

had these four conserved nucleotides plus one additional position adjacent to the internal loop in helix III (G28) randomized, and so allowed explore whether or not any combinations of mutations would allow for restoration of activity. In library 3, only the non-conserved positions in the internal loop were randomized. In library 2, the non-conserved positions and the adjacent base-pairs in the helices II and III were randomized. In library 1 all of the internal loop plus the adjacent base-pairs of helix II and III were randomized. For all of these libraries the sequence of active clones was the same as wild-type except at position 28. The variability allowed at position 28 is consistent with the single-mutation studies that showed this position could be any base and still retain the full activity of the wild-type ribozyme. The conservation of the entire loop sequence in these libraries shows that, for the basic architecture and size of the 49mer ribozyme, a catalytic solution near the optimum has been found for both the *in cis* and *in trans* reactions.

It is possible there are other solutions requiring a complete rearrangement of the ribozyme that would give more efficient catalysis. If so, then the type of saturation mutagenesis libraries used for this study would be unlikely to find them. Buskirk and colleagues have shown that the chance of finding a functional RNA structure is related to its size when selecting for RNA transcription factors. Active transcription factors were found in libraries with larger randomized regions (N_{40}) compared to smaller ones (N_{10}) (106). It is possible that the size of the starting motif is too small to allow for the exploration of other potential catalysts similar to the original without destroying the original structure.

Second, assuming that a major rearrangement of structure would be required to facilitate better catalytic activity, much more of the ribozyme would need to be randomized. It has been theorized that a mutation rate of at least ~20% at each position is required to alter the secondary structure of RNA (107), and this has been supported by experimental evidence with both aptamers (108) and ribozymes (69). By only randomizing the internal loop region of the 49mer, there was little opportunity for new secondary structure to be explored.

Even though selections of these libraries were not successful at finding a novel or better catalyst, they acted as a positive control for the method in general. The selections demonstrated that IVC is a useful method of evolving ribozymes by showing that genes encoding active ribozymes could be selected from a complex mixture of similar sequences.

157mer selections

To compare IVC with SELEX we used IVC to select a library that had previously been selected using SELEX to yield Diels-Alderase ribozymes (3). The SELEX experiment started from a library of $\sim 2 \times 10^{14}$ RNA molecules, each 157 bases long with the central 120 nucleotides randomized. To select a similar diversity library using IVC would require a 1.1 litre emulsion. Although this is not impossible, we chose an alternative strategy. The IVC selections were started with cDNA from the 157mer library purified after the 5th round of SELEX (3). This has the added advantage that SELEX is based on selection for intramolecular, single-turnover reactions and, because of this, it can very easily select ribozymes which give even the most modest rate acceleration (see (22)).

Hence, the starting library should already be enriched for active ribozymes. Indeed, this library showed low, but detectable, Diels-Alderase activity even before selection (Figure 23a).

The library was selected using IVC as described in Figure 4 under both single- and multiple-turnover conditions. Diels-Alderase activity increased more quickly under single-turnover than under multiple-turnover conditions, consistent with the theoretical predictions of the enrichment rates under these two regimes (see above). By round six of the single-turnover IVC selections the activity of the pool was equal to that of the 49mer ribozyme (85), while the activity of the pool from the multiple-turnover selections was about half the activity of the 49mer by round nine (Figure 19). At this point, individual clones were analysed and those with the highest activity were sequenced. Hence, under the single-turnover regime, the library had undergone a total of 11 rounds of selection (5 by SELEX then 6 by IVC), exactly the same number of rounds of selection as used to isolate ribozymes from the same library using SELEX (3).

All ribozymes isolated from the SELEX library after 11 rounds of selection shared the same small secondary structural motif consisting of 3 helices, an asymmetric internal loop, and the formally single-stranded 5'-end (Figure 20) (3). Nucleotides of the 5' terminus are thought to base-pair with nucleotides on both sides of the internal loop forming a pseudoknot fold (99). The crystal structure of the 49mer ribozyme has recently been solved, confirming this structural prediction (Serganov et al., 2005).

Using IVC nine distinct sequence classes were identified (Figure 23). Classes 1-4 were found under both single-turnover (ST) and multiple-turnover (MT) selection regimes, classes 5,6 and 9 were only found under ST conditions, and classes 6 and 8 were only seen under MT conditions. Clones from Classes 1-7 contain the same secondary structure motif found in the SELEX ribozymes, but differ from each other in the location of the internal loop and the lengths of the helices (Figure 19b). The most common class of ribozymes from both ST and MT IVC selections (class 1) was identical to the most common class identified using SELEX (group 2). Indeed, IVC classes 1, 3, and 7 are the same as SELEX groups 2, 7 and 5 respectively, but the other IVC classes have not been observed before.

Interestingly, the IVC selected ribozymes show some striking differences in the frequency of mutations within the internal loop compared to SELEX ribozymes with the same structural motif (see Figure 19c). U12, is found in ~25% of all ribozymes selected by IVC, but is not seen in any of the ribozymes from SELEX. Similarly, G15 is found in nearly half of the ribozymes selected by IVC and in 75% of class 1 ribozymes, but this sequence was seen in only a single ribozyme from SELEX. This suggests that there is a selective advantage for these sequences using IVC that is not present using SELEX.

Since, in the single-turnover selections, we were selecting for activity in trans at a very low diene substrate concentration (9 nM, assuming 70 nm diameter droplets), there should be selective pressure to reduce the K_m . The most common sequence under the single-turnover IVC regime, is a class 1 ribozyme

with G15 in the internal loop. Assayed for the ability to catalyse the in trans reaction under multiple-turnover conditions (Table 3) this ribozyme, represented by clone E12, does show a modest 2.5-fold reduction in K_m for the diene compared to ribozyme F12, which has the same internal loop sequence as the 49mer ribozyme (with A15 in the internal loop). However, this is accompanied by a 4.8-fold increase in K_m for the dienophile and a 2.6-fold decrease in k_{cat} which makes the overall catalytic efficiency ($k_{cat}/K_m(\text{diene})K_m(\text{dienophile})$) 7.1-fold lower and therefore this clone would not be expected to have any particular selective advantage. Indeed the theoretical enrichment in each round (Equation 6) would only be 1.72-fold for the E12 mutant.

The second most common class of ribozymes from the IVC study (class 2), which constitutes 25% of the ribozymes analysed has an A12U mutation, which causes a potential U12:A15 base-pair that would make the internal loop smaller. In two of the class 2 sequences the length of transcribed RNA is only 106 nt instead of 156 nt for the rest of the class. The 50 bp truncation has removed much of the 3' end of the gene with the 3' primer-binding site placed immediately after the helix II motif, and was allowed to accumulate since there was no size-selection between selection rounds. The structural motifs of this class are already clustered near the 5' end of the sequence with shorter helices and terminal loops than some of the other groups (Figure 19b). This deletion shows how selective pressure – either for increased activity or because of a PCR advantage for the shorter sequence – is beginning to reduce the RNA to a size that is more similar to the minimal 49mer motif that was engineered using rational design (3).

A class 2 ribozyme from the IVC selections (clone G11) was also assayed for the ability to catalyse the *in trans* reaction under multiple-turnover conditions (Table 1). Its catalytic efficiency ($k_{\text{cat}}/K_{\text{m(diene)}}K_{\text{m(dienophile)}}$) was ~27-times lower than either the 49mer ribozyme, or ribozyme F12, which have identical internal loop sequences. The theoretical enrichment for this sequence (Equation 6) would only be ~1.2-fold per round under either single- or multiple-turnover conditions. Hence, it is a little surprising that such ribozymes should be so common in the IVC selections.

Mutational analysis of the internal loop region of the 49mer ribozyme (99) has shown that mutants with either the A15G (like clone E12) or A12U mutation (like clone G11) are 10-fold and 20-fold less active respectively than wild-type when assayed using a tethered anthracene substrate at a single maleimide concentration (in *cis* reaction). Interestingly the same mutational analysis showed that the A12G mutation has nearly the same activity with a tethered substrate as the wild-type sequence, but even so, the mutation was rare in both studies.

It could be that these subtle differences in relative rates when selecting for the *in cis* reaction (SELEX) and in *trans* reaction (IVC) account for the different frequencies of these sequences in the SELEX and IVC ribozymes (Figure 19c).

Perhaps the most interesting result of the IVC selections is that primary sequence analysis and secondary structure predictions indicate that the class 9 (from the single-turnover selections) and class 8 (from the multiple-turnover selections) ribozymes have no structural or sequence homology to the catalytic

motif found in all the ribozymes isolated with SELEX. The secondary structure of clone E03, a class 9 ribozyme, was confirmed by structural probing experiments indicating that this class of Diels-Alderase ribozymes indeed has a completely novel fold (Figure 20).

Clone E03, was shown to be an active Diels-Alderase ribozyme capable of catalysing the *in trans* reaction using diene and dienophile free in solution under multiple-turnover conditions. The kinetic properties were quite similar to the wild-type 49mer and ribozyme F12, which has the same internal loop sequence. From this, it appears that IVC has found a novel solution for Diels-Alder catalysis. The overall length and primer binding sequences are unchanged, indicating that there was not a simple advantage in PCR efficiency that would allow this sequence to be preferentially amplified. There was also no difference in yield of RNA from *in vitro* transcription in bulk solution. The kinetic parameters in bulk solution are if anything slightly worse for the class 8 fold compared to ribozyme F12 or the 49mer ribozyme, which share the same internal loop sequence: overall catalytic efficiency ($k_{cat}/K_m(\text{diene})K_m(\text{dienophile})$) is 6.4-fold lower, and the theoretical enrichment factor (Equation 6) would be ~1.9-fold. There may be something about the micro environment within an emulsion that would allow it to be selected. For example, there may be some interaction with at the microdroplet interface. There may also simply be less of a selective disadvantage for these ribozymes when selected for catalysis of the *in trans* reaction (IVC) compared to the *in cis* reaction (SELEX). For example, the transcription rate of the E03 ribozyme may be better when emulsified - although there were not gross differences observed in bulk solution (data not shown).

Alternatively, the E03 ribozyme may fold into an active configuration with higher efficiency *in situ* than the F12-like sequences.

Perhaps the most striking feature of the kinetic analysis of the Diels-Alderase ribozymes selected by IVC is how similar are the kinetic parameters for all ribozymes tested. Notably, the ribozyme with the novel fold (E03) has a $K_{m(\text{diene})}$, $K_{m(\text{dienophile})}$ and k_{cat} values which are very close to those of both the 49mer and clone F12, which share the same internal loop sequence.

It is tempting to speculate that this is not purely a coincidence, but is related fundamentally to the mechanism of the Diels-Alder reaction. All the ribozymes are strongly product inhibited with almost identical K_i (product) of $\sim 50 \mu\text{M}$, which is lower than the K_m for either substrate. This is almost certainly related to the fact that the ribozyme must bind the transition state, but the transition state of the Diels-Alder reaction is very close to that of the reaction product. However, there is kinetic evidence to suggest that non-covalent catalysis by the 49mer ribozyme, and by a range of other disparate non-covalent catalysts of the Diels-Alder reaction (including several catalytic antibodies, cyclodextrin, and Rebek's tennis ball (84)). None of the catalysts achieve significant specific binding of transition states that is the hallmark of enzyme catalysis (84). Indeed, the 49mer ribozyme is thought to have a preformed catalytic pocket as probing data show no major structural changes on substrate or product binding (99). The same lack of significant structural changes on product binding was observed for the ribozymes selected using IVC (data not shown). Rate acceleration is thought to arise predominantly from binding of reactants, converting a second-order

reaction of diene with dienophile into a first-order reaction of the termolecular complex of catalyst, diene and dienophile (84). This is catalysis by an “entropy trap” in the sense of “overcoming the unfavorable entropy of activation” (109), or, given the current, more restricted definition of an entropy trap, catalysis by “approximation”, i.e. “the bringing together of two or more reactants in the active site” (110).

Lewis acid catalysis has been proposed to contribute significantly to rate acceleration by pyridine-modified Diels-Alderase ribozymes selected using SELEX, as they show an absolute dependence on Cu^{2+} or Ni^{2+} (68,111). These ribozymes only catalyse the *in cis* reaction and have therefore been assayed under single-turnover conditions. The 9 ribozymes analysed displayed k_{cat} and K_{m} values that spanned 74 and 80-fold ranges respectively. However, their values roughly compensate for one another, as the values of $k_{\text{cat}}/K_{\text{m}}$ varied by only 5-fold. Further, there was a general trend for product inhibition to correlate with k_{cat} , consistent with the product-like nature of the Diels-Alder transition state. Further mutagenesis of one of these ribozymes (DA22) and selection by SELEX yielded ribozymes with up to 45-fold improvements in catalytic efficiency ($k_{\text{cat}}/K_{\text{m}}$), primarily due to a decrease in K_{m} . One of the selected ribozymes also acquired the ability to function with Ni^{2+} as the only transition-metal ion (69).

The results from this and earlier studies of Diels-Alderase could go some way to explaining why nature seems to prefer other mechanisms of C-C bond formation, such as the aldol condensation reaction, over the Diels-Alder cycloaddition. The similarity between the Diels-Alder transition state and the

product creates a problem for the catalyst: stabilisation of the transition state resulting in an almost inevitable increase in affinity for the product, reducing the rate of product release and causing product inhibition. Indeed, in the most efficient artificial Diels-Alderase described to date, the catalytic antibody 1E9, uncatalysed SO₂ elimination from the Diels-Alder adduct is programmed so as to avoid product inhibition (112). Similarly, the most well characterised natural Diels-Alderase, macrophomate synthase, catalyses the elimination CO₂ to prevent product inhibition (83). Macrophomate synthase and the two other examples of natural Diels-Alderases, solanapyrone synthase and lovastatin nonaketide synthase, also catalyse the conversion of the corresponding substrates into reactive Diels-Alder substrates that are not released from the active sites (82). They are thus, all producers of reactive substrates with an entropy trap for spontaneous [4+2] cycloaddition. This strategy has the obvious advantage that it allows the use of a highly reactive substrate that is not stable in reaction medium.

A requirement for enzymes to evolve to catalyse, not only the Diels-Alder cycloaddition, but also a second reaction in which the cycloadduct is the substrate, may present a significant evolutionary barrier. This barrier may be even higher if it is also necessary to catalyse a reaction to create a reactive substrate for the [4 + 2] cycloaddition.

As can be seen from Figure 28, none of the natural or artificial Diels-Alderases which have been characterised kinetically have k_{cat} of $\geq 1 \text{ s}^{-1}$, $k_{cat}/K_m(\text{diene}) \geq 10^4 \text{ s}^{-1}$, $k_{cat}/K_m(\text{dienophile}) \geq 10^3 \text{ s}^{-1}$, or $k_{cat}/K_m(\text{diene})K_m(\text{dienophile}) \geq 10^6 \text{ s}^{-1}$. It may be difficult

for Diels-Alderases to pass this upper limit due to the nature of the Diels-Alder reaction itself.

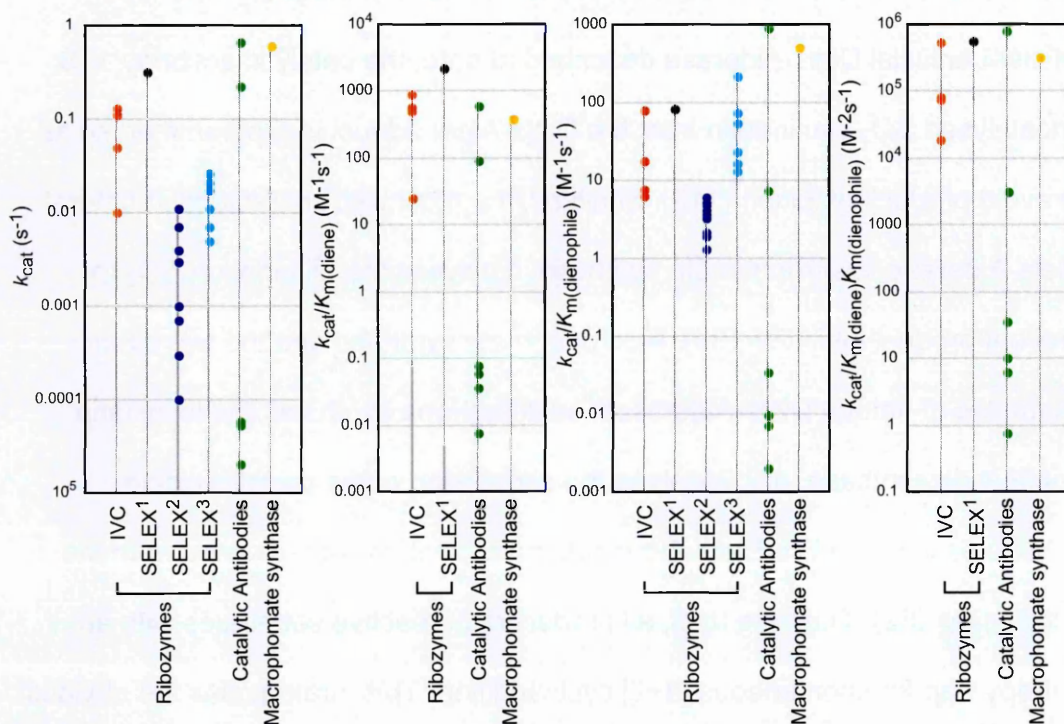


Figure 28 Kinetic parameters of natural and artificial Diels-Alderases.

The k_{cat} , $k_{cat}/K_m(diene)$, $k_{cat}/K_m(dienophile)$ and $k_{cat}/K_m(diene)K_m(dienophile)$ for four ribozymes selected by IVC (red circles), the 49mer ribozyme derived by SELEX (SELEX¹) (85) (black circle), nine ribozymes isolated from a different SELEX experiment (SELEX²) [tarasow, 1997; tarasow, 2000] (blue circles), six ribozymes derived by mutation and SELEX of a ribozyme from the previous study (SELEX³) [(69)(cyan circles), six catalytic antibodies, 1E9, 39A11, 13G5, 4D5, 22C8 and 7D4 (data summarised in 84) (green circles) and macrophomate synthase (83) (orange circle). Data for the SELEX² and SELEX³ ribozymes was determined under single-turnover conditions, and it was not possible to determine the $K_m(diene)$ as these ribozymes were only active when the diene was tethered to the RNA. Circles for enzymes with very similar kinetic parameters are overlaid.

The Diels-Alderase ribozymes selected by IVC have kinetic constants which already approach this upper limit (Figure 28) and hence it may be difficult to select Diels-Alderase ribozyme that are significantly better than those described here, even by selecting for catalysis of the *in trans* reaction under multiple-turnover conditions using IVC.

However, with other systems which do not have such mechanistic limitations it may be possible to use IVC to select ribozymes that are very fast and efficient. There is a view from some enzymologists that ribozymes (and deoxyribozymes) may be inherently incapable of generating the same rate enhancements as those achieved by protein enzymes due to the lower chemical complexity of nucleic acids. Others, however, feel that this is not necessarily the case. For example, it has been argued that, by using three or four catalytic strategies in parallel a ribozyme or deoxyribozyme should theoretically be able to catalyse RNA cleavage by internal phosphoester transfer at rates at least equivalent to the protein enzyme Ribonuclease A (a nearly 10^{13} -fold rate acceleration) (87).

However, attempts to improve ribozymes by *in vitro* selection using SELEX, which is based on selecting for intramolecular single turnover reactions, are likely to fail (see for example 105) as there is neither selection for substrate recognition nor high k_{cat} . Using IVC, however, it may prove possible to select for more efficient ribozymes, if they exist, and thereby answer some important questions about the true catalytic potential of RNA. Ribozymes catalysing RNA cleavage such as the X-motif and MR8 ribozyme (113,114) may make excellent candidates for improvement by *in vitro* selection using IVC.

References

1. Gesteland, R. F., Cech, T., and Atkins, J. F. (1999) *The RNA world : the nature of modern RNA suggests a prebiotic RNA*, 2nd Ed. Cold Spring Harbor monograph series, 37, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
2. Tawfik, D. S., and Griffiths, A. D. (1998) Man-made cell-like compartments for molecular evolution *Nat Biotechnol* **16**, 652-656
3. Seelig, B., and Jaschke, A. (1999) A small catalytic RNA motif with Diels-Alderase activity *Chem Biol* **6**, 167-176
4. Perelson, A. S., and Oster, G. F. (1979) Theoretical studies of clonal selection: minimal antibody repertoire size and reliability of self-non-self discrimination *J Theor Biol* **81**, 645-670
5. Varga, J. M., Kalchschmid, G., Klein, G. F., and Fritsch, P. (1991) Mechanism of allergic cross-reactions--I. Multispecific binding of ligands to a mouse monoclonal anti-DNP IgE antibody *Mol Immunol* **28**, 641-654
6. Lancet, D., Sadovsky, E., and Seidemann, E. (1993) Probability model for molecular recognition in biological receptor repertoires: significance to the olfactory system *Proc Natl Acad Sci U S A* **90**, 3715-3719
7. Wilson, D. S., and Szostak, J. W. (1999) In vitro selection of functional nucleic acids *Annu Rev Biochem* **68**, 611-647
8. Zaccolo, M., Williams, D. M., Brown, D. M., and Gherardi, E. (1996) An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues *J Mol Biol* **255**, 589-603
9. Kayushin, A. L., Korosteleva, M. D., Miroshnikov, A. I., Kosch, W., Zubov, D., and Piel, N. (1996) A convenient approach to the synthesis of trinucleotide phosphoramidites--synthons for the generation of oligonucleotide/peptide libraries *Nucleic Acids Res* **24**, 3748-3755
10. Stemmer, W. P. (1994) Rapid evolution of a protein in vitro by DNA shuffling *Nature* **370**, 389-391
11. Zhao, H., Giver, L., Shao, Z., Affholter, J. A., and Arnold, F. H. (1998) Molecular evolution by staggered extension process (StEP) in vitro recombination *Nat Biotechnol* **16**, 258-261
12. Crameri, A., Raillard, S. A., Bermudez, E., and Stemmer, W. P. (1998) DNA shuffling of a family of genes from diverse species accelerates directed evolution *Nature* **391**, 288-291
13. Minshull, J., and Stemmer, W. P. (1999) Protein evolution by molecular breeding *Curr Opin Chem Biol* **3**, 284-290
14. Ness, J. E., Welch, M., Giver, L., Bueno, M., Cherry, J. R., Borchert, T. V., Stemmer, W. P., and Minshull, J. (1999) DNA shuffling of subgenomic sequences of subtilisin *Nat Biotechnol* **17**, 893-896
15. Babbitt, P. C., and Gerlt, J. A. (1997) Understanding enzyme superfamilies. Chemistry As the fundamental determinant in the evolution of new catalytic activities *J Biol Chem* **272**, 30591-30594
16. Bittker, J. A., Le, B. V., and Liu, D. R. (2002) Nucleic acid evolution and minimization by nonhomologous random recombination *Nat Biotechnol* **20**, 1024-1029

17. Bittker, J. A., Le, B. V., Liu, J. M., and Liu, D. R. (2004) Directed evolution of protein enzymes using nonhomologous random recombination *Proc Natl Acad Sci U S A* **101**, 7011-7016
18. Amann, R. I., Ludwig, W., and Schleifer, K. H. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation *Microbiol Rev* **59**, 143-169
19. Schloss, P. D., and Handelsman, J. (2003) Biotechnological prospects from metagenomics *Curr Opin Biotechnol* **14**, 303-310
20. Brennan, Y., Callen, W. N., Christoffersen, L., Dupree, P., Goubet, F., Healey, S., Hernandez, M., Keller, M., Li, K., Palackal, N., Sittenfeld, A., Tamayo, G., Wells, S., Hazlewood, G. P., Mathur, E. J., Short, J. M., Robertson, D. E., and Steer, B. A. (2004) Unusual microbial xylanases from insect guts *Appl Environ Microbiol* **70**, 3609-3617
21. Palackal, N., Brennan, Y., Callen, W. N., Dupree, P., Frey, G., Goubet, F., Hazlewood, G. P., Healey, S., Kang, Y. E., Kretz, K. A., Lee, E., Tan, X., Tomlinson, G. L., Verruto, J., Wong, V. W., Mathur, E. J., Short, J. M., Robertson, D. E., and Steer, B. A. (2004) An evolutionary route to xylanase process fitness *Protein Sci* **13**, 494-503
22. Griffiths, A. D., and Tawfik, D. S. (2000) Man-made enzymes-from design to in vitro compartmentalisation *Curr Opin Biotechnol* **11**, 338-353.
23. Amstutz, P., Forrer, P., Zahnd, C., and Pluckthun, A. (2001) In vitro display technologies: novel developments and applications *Curr Opin Biotechnol* **12**, 400-405.
24. Georgiou, G., Stathopoulos, C., Daugherty, P. S., Nayak, A. R., Iverson, B. L., and Curtiss, R., 3rd. (1997) Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines *Nat Biotechnol* **15**, 29-34
25. Griffiths, A. D., and Duncan, A. R. (1998) Strategies for selection of antibodies by phage display *Curr Opin Biotechnol* **9**, 102-108
26. Keefe, A. D., and Szostak, J. W. (2001) Functional proteins from a random-sequence library *Nature* **410**, 715-718.
27. Pluckthun, A., Schaffitzel, C., Hanes, J., and Jermutus, L. (2000) In vitro selection and evolution of proteins *Adv Protein Chem* **55**, 367-403
28. Schatz, P. J., Cull, M. G., Martin, E. L., and Gates, C. M. (1996) Screening of peptide libraries linked to lac repressor *Methods Enzymol* **267**, 171-191
29. Sidhu, S. S. (2000) Phage display in pharmaceutical biotechnology *Curr Opin Biotechnol* **11**, 610-616.
30. Wittrup, K. D. (2001) Protein engineering by cell-surface display *Curr Opin Biotechnol* **12**, 395-399.
31. Odegrip, R., Coomber, D., Eldridge, B., Hederer, R., Kuhlman, P. A., Ullman, C., FitzGerald, K., and McGregor, D. (2004) CIS display: In vitro selection of peptides from libraries of protein-DNA complexes *Proc Natl Acad Sci U S A* **101**, 2806-2810
32. Yonezawa, M., Doi, N., Kawahashi, Y., Higashinakagawa, T., and Yanagawa, H. (2003) DNA display for in vitro selection of diverse peptide libraries *Nucleic Acids Res* **31**, e118
33. Yonezawa, M., Doi, N., Higashinakagawa, T., and Yanagawa, H. (2004) DNA display of biologically active proteins for in vitro protein selection *J Biochem (Tokyo)* **135**, 285-288
34. Sepp, A., Tawfik, D. S., and Griffiths, A. D. (2002) Microbead display by in vitro compartmentalisation: selection for binding using flow cytometry *FEBS Lett* **532**, 455-458
35. Fujita, S., Taki, T., and Taira, K. (2005) Selection of an Active Enzyme by Phage Display on the Basis of the Enzyme's Catalytic Activity in vivo *Chembiochem* **6**, 315-321
36. Tuerk, C., and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase *Science* **249**, 505-510
37. Bartel, D. P., and Szostak, J. W. (1993) Isolation of new ribozymes from a large pool of random sequences [see comment] *Science* **261**, 1411-1418
38. Wiegand, T. W., Janssen, R. C., and Eaton, B. E. (1997) Selection of RNA amide synthases *Chem Biol* **4**, 675-683

39. Sengle, G., Eisenfuhr, A., Arora, P. S., Nowick, J. S., and Famulok, M. (2001) Novel RNA catalysts for the Michael reaction *Chem Biol* **8**, 459-473
40. Tsukiji, S., Pattnaik, S. B., and Suga, H. (2003) An alcohol dehydrogenase ribozyme *Nat Struct Biol* **10**, 713-717
41. Kumar, R. K., and Yarus, M. (2001) RNA-catalyzed amino acid activation *Biochemistry* **40**, 6998-7004
42. Doi, N., and Yanagawa, H. (1999) STABLE: protein-DNA fusion system for screening of combinatorial protein libraries in vitro *FEBS Lett* **457**, 227-230
43. Lee, Y. F., Tawfik, D. S., and Griffiths, A. D. (2002) Investigating the target recognition of DNA cytosine-5 methyltransferase HhaI by library selection using in vitro compartmentalisation *Nucleic Acids Res* **30**, 4937-4944
44. Cohen, H. M., Tawfik, D. S., and Griffiths, A. D. (2004) Altering the sequence specificity of HaeIII methyltransferase by directed evolution using in vitro compartmentalization *Protein Eng Des Sel* **17**, 3-11
45. Ghadessy, F. J., Ong, J. L., and Holliger, P. (2001) Directed evolution of polymerase function by compartmentalized self-replication *Proc Natl Acad Sci U S A* **98**, 4552-4557.
46. Ghadessy, F. J., Ramsay, N., Boudsocq, F., Loakes, D., Brown, A., Iwai, S., Vaisman, A., Woodgate, R., and Holliger, P. (2004) Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution *Nat Biotechnol* **22**, 755-759
47. Griffiths, A. D., and Tawfik, D. S. (2003) Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization *Embo J* **22**, 24-35
48. Bernath, K., Hai, M., Mastrobattista, E., Griffiths, A. D., Magdassi, S., and Tawfik, D. S. (2004) In vitro compartmentalization by double emulsions: sorting and gene enrichment by fluorescence activated cell sorting *Anal Biochem* **325**, 151-157
49. Bernath, K., Magdassi, S., and Tawfik, D. S. (2005) Directed evolution of protein inhibitors of DNA-nucleases by in vitro compartmentalization (IVC) and nano-droplet delivery *J Mol Biol* **345**, 1015-1026
50. Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., and Cech, T. R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena *Cell* **31**, 147-157
51. Peebles, C. L., Perlman, P. S., Mecklenburg, K. L., Petrillo, M. L., Tabor, J. H., Jarrell, K. A., and Cheng, H. L. (1986) A self-splicing RNA excises an intron lariat *Cell* **44**, 213-223
52. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme *Cell* **35**, 849-857
53. Buzayan, J. M., Hampel, A., and Bruening, G. (1986) Nucleotide sequence and newly formed phosphodiester bond of spontaneously ligated satellite tobacco ringspot virus RNA *Nucleic Acids Res* **14**, 9729-9743
54. Sharmeen, L., Kuo, M. Y., Dinter-Gottlieb, G., and Taylor, J. (1988) Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage *J Virol* **62**, 2674-2679
55. Saville, B. J., and Collins, R. A. (1990) A site-specific self-cleavage reaction performed by a novel RNA in *Neurospora* mitochondria *Cell* **61**, 685-696
56. Nissen, P., Hansen, J., Ban, N., Moore, P. B., and Steitz, T. A. (2000) The structural basis of ribosome activity in peptide bond synthesis *Science* **289**, 920-930
57. Valadkhan, S., and Manley, J. L. (2001) Splicing-related catalysis by protein-free snRNAs *Nature* **413**, 701-707
58. Joyce, G. F. (2004) DIRECTED EVOLUTION OF NUCLEIC ACID ENZYMES *Annual Review of Biochemistry* **73**, 791-836
59. Cech, T. R. (2002) Ribozymes, the first 20 years *Biochem Soc Trans* **30**, 1162-1166
60. Wecker, M., Smith, D., and Gold, L. (1996) In vitro selection of a novel catalytic RNA: characterization of a sulfur alkylation reaction and interaction with a small peptide *Rna* **2**, 982-994
61. Wilson, C., and Szostak, J. W. (1995) In vitro evolution of a self-alkylating ribozyme *Nature* **374**, 777-782

-
62. Illangasekare, M., Sanchez, G., Nickles, T., and Yarus, M. (1995) Aminoacyl-RNA synthesis catalyzed by an RNA *Science* **267**, 643-647
 63. Lohse, P. A., and Szostak, J. W. (1996) Ribozyme-catalysed amino-acid transfer reactions *Nature* **381**, 442-444
 64. Illangasekare, M., and Yarus, M. (1997) Small-molecule-substrate interactions with a self-aminoacylating ribozyme *J Mol Biol* **268**, 631-639
 65. Suga, H., Lohse, P. A., and Szostak, J. W. (1998) Structural and kinetic characterization of an acyl transferase ribozyme *J Am Chem Soc* **120**, 1151-1156
 66. Huang, F., Bugg, C. W., and Yarus, M. (2000) RNA-Catalyzed CoA, NAD, and FAD synthesis from phosphopantetheine, NMN, and FMN *Biochemistry* **39**, 15548-15555
 67. Jadhav, V. R., and Yarus, M. (2002) Acyl-CoAs from coenzyme ribozymes *Biochemistry* **41**, 723-729
 68. Tarasow, T. M., Tarasow, S. L., and Eaton, B. E. (1997) RNA-catalysed carbon-carbon bond formation *Nature* **389**, 54-57
 69. Tarasow, T. M., Kellogg, E., Holley, B. L., Nieuwlandt, D., Tarasow, S. L., and Eaton, B. E. (2004) The effect of mutation on RNA Diels-Alderase *J Am Chem Soc* **126**, 11843-11851
 70. Houk, K. N., Gonzalez, J., and Li, Y. (1995) Pericyclic Reaction Transition-States - Passions and Punctilios, 1935-1995 *Accounts of Chemical Research* **28**, 81-90
 71. Houk, K. N., Lin, Y. T., and Brown, F. K. (1986) Evidence for the Concerted Mechanism of the Diels-Alder Reaction of Butadiene with Ethylene *Journal of the American Chemical Society* **108**, 554-556
 72. Tarasow, T. M., and Eaton, B. E. (1999) The Diels-Alder reaction and biopolymer catalysis *Cell Mol Life Sci* **55**, 1463-1472
 73. Arigoni, D. (2000) Comment concerning the review article: The Diels-Alder reaction and biopolymer catalysis by T.M. Tarasow and B.E. Eaton *Cell Mol Life Sci* **57**, 982; discussion 983-984
 74. Jorgensen, W. L., Lim, D. C., and Blake, J. F. (1993) Abinitio Study of Diels-Alder Reactions of Cyclopentadiene with Ethylene, Isoprene, Cyclopentadiene, Acrylonitrile, and Methyl Vinyl Ketone *Journal of the American Chemical Society* **115**, 2936-2942
 75. Grieco, P. A., Garner, P., Yoshida, K., and Huffman, J. C. (1983) Aqueous Intermolecular Diels-Alder Chemistry - Novel Products Derived from Substituted Benzoquinone-Diene Carboxylate Adducts Via Tandem Michael Reactions *Tetrahedron Letters* **24**, 3807-3810
 76. Breslow, R., Maitra, U., and Rideout, D. (1983) Selective Diels-Alder Reactions in Aqueous-Solutions and Suspensions *Tetrahedron Letters* **24**, 1901-1904
 77. Blake, J. F., Lim, D., and Jorgensen, W. L. (1994) Enhanced Hydrogen-Bonding of Water to Diels-Alder Transition- States - Ab-Initio Evidence *Journal of Organic Chemistry* **59**, 803-805
 78. Birney, D. M., and Houk, K. N. (1990) Transition Structures of the Lewis Acid-Catalyzed Diels-Alder Reaction of Butadiene with Acrolein - the Origins of Selectivity *Journal of the American Chemical Society* **112**, 4127-4133
 79. Corey, E. J., Loh, T. P., Roper, T. D., Azimioara, M. D., and Noe, M. C. (1992) The Origin of Greater Than 200/1 Enantioselectivity in a Catalytic Diels-Alder Reaction as Revealed by Physical and Chemical Studies *Journal of the American Chemical Society* **114**, 8290-8292
 80. Laschat, S. (1996) Pericyclic reaction in biological systems - does nature know about the Diels-Alder reaction? *Angew Chem Int Ed Engl* **35**, 289
 81. Stocking, E. M., and Williams, R. M. (2003) Chemistry and biology of biosynthetic Diels-Alder reactions *Angew Chem Int Ed Engl* **42**, 3078-3115
 82. Pohnert, G. (2001) Diels-Alderase *ChemBiochem* **2**, 873-875
 83. Ose, T., Watanabe, K., Mie, T., Honma, M., Watanabe, H., Yao, M., Oikawa, H., and Tanaka, I. (2003) Insight into a natural Diels-Alder reaction from the structure of macrophomate synthase *Nature* **422**, 185-189
-

84. Kim, S. P., Leach, A. G., and Houk, K. N. (2002) The origins of noncovalent catalysis of intermolecular Diels-Alder reactions by cyclodextrins, self-assembling capsules, antibodies, and RNAses *J Org Chem* **67**, 4250-4260
85. Seelig, B., Keiper, S., Stuhlmann, F., and Jaschke, A. (2000) Enantioselective Ribozyme Catalysis of a Bimolecular Cycloaddition Reaction *Angew Chem Int Ed Engl* **39**, 4576-4579
86. Kumar, C. V., Punzalan, E. H. A., and Tan, W. B. (2000) Adenine-thymine base pair recognition by an anthryl probe from the DNA minor groove *Tetrahedron* **56**, 7027-7040
87. Emilsson, G. M., Nakamura, S., Roth, A., and Breaker, R. R. (2003) Ribozyme speed limits *Rna* **9**, 907-918
88. Ellinger, T., and Ehricht, R. (1998) Single-step purification of T7 RNA polymerase with a 6-histidine tag *Biotechniques* **24**, 718-720
89. Soukup, G. A., and Breaker, R. R. (1999) Relationship between internucleotide linkage geometry and the stability of RNA *Rna* **5**, 1308-1325
90. Cleland, W. W. (1970) *The Enzymes* **2**, 2-65
91. Dalziel, K. (1970) *Enzymes* **11**, 2-60
92. Mueller, J. D., Putz, B., and Hofler, H. (1997) Self-sustained sequence replication (3SR): an alternative to PCR *Histochem Cell Biol* **108**, 431-437
93. Gingeras, T. R., Whitfield, K. M., and Kwoh, D. Y. (1990) Unique features of the self-sustained sequence replication (3SR) reaction in the in vitro amplification of nucleic acids *Ann Biol Clin (Paris)* **48**, 498-501
94. Fahy, E., Kwoh, D. Y., and Gingeras, T. R. (1991) Self-sustained sequence replication (3SR): an isothermal transcription-based amplification system alternative to PCR *Methods Appl* **1**, 25-33
95. Ke, A., and Doudna, J. A. (2004) Crystallization of RNA and RNA-protein complexes *Methods* **34**, 408-414
96. Price, S. R., Ito, N., Oubridge, C., Avis, J. M., and Nagai, K. (1995) Crystallization of RNA-protein complexes. I. Methods for the large-scale preparation of RNA suitable for crystallographic studies *J Mol Biol* **249**, 398-408
97. Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans* *Nature* **391**, 806-811
98. Milburn, S. C., Goldrick, M., and Winkler, M. (2003), pp. 21, Ambion, Inc., USA
99. Keiper, S., Bebenroth, D., Seelig, B., Westhof, E., and Jaschke, A. (2004) Architecture of a Diels-Alderase Ribozyme with a Preformed Catalytic Pocket *Chemistry & Biology* **11**, 1217-1227
100. Schneider, T. D., and Stephens, R. M. (1990) Sequence logos: a new way to display consensus sequences *Nucleic Acids Res* **18**, 6097-6100
101. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction *Nucleic Acids Res* **31**, 3406-3415
102. Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure *J Mol Biol* **288**, 911-940
103. Isel, C., Marquet, R., Keith, G., Ehresmann, C., and Ehresmann, B. (1993) Modified nucleotides of tRNA(3Lys) modulate primer/template loop-loop interaction in the initiation complex of HIV-1 reverse transcription *J Biol Chem* **268**, 25269-25272
104. Stuhlmann, F., and Jaschke, A. (2002) Characterization of an RNA active site: interactions between a Diels-Alderase ribozyme and its substrates and products *J Am Chem Soc* **124**, 3238-3244
105. Santoro, S. W., and Joyce, G. F. (1997) A general purpose RNA-cleaving DNA enzyme *Proc Natl Acad Sci U S A* **94**, 4262-4266
106. Buskirk, A. R., Kehayova, P. D., Landrigan, A., and Liu, D. R. (2003) In vivo evolution of an RNA-based transcriptional activator *Chem Biol* **10**, 533-540

-
107. Huynen, M. A. (1996) Exploring phenotype space through neutral evolution *J Mol Evol* **43**, 165-169
 108. Huang, Z., and Szostak, J. W. (2003) Evolution of aptamers with a new specificity and new secondary structures from an ATP aptamer *Rna* **9**, 1456-1463
 109. Westheimer, F. H. (1962) Mechanisms related to enzyme catalysis *Adv Enzyme Regul* **24**, 441-482
 110. Westheimer, F. H. (1985) The Discovery of the Mechanisms of Enzyme Action, 1947-1963 *Advances in Physical Organic Chemistry* **21**, 1-35
 111. Tarasow, T. M., Tarasow, S. L., and Eaton, B. E. (2000) RNA Diels-Alderase: Relationships between Unique Sequences and Catalytic Function *J Am Chem Soc* **122**, 1015-1021
 112. Hilvert, D., and Hill, K. W. (1991) Antibody catalysis of concerted, carbon-carbon bond-forming reactions *Methods Enzymol* **203**, 352-369
 113. Tang, J., and Breaker, R. R. (2000) Structural diversity of self-cleaving ribozymes *Proc Natl Acad Sci U S A* **97**, 5784-5789
 114. Lazarev, D., Puskarz, I., and Breaker, R. R. (2003) Substrate specificity and reaction kinetics of an X-motif ribozyme *Rna* **9**, 688-697